

UNIVERSIDAD CARLOS III DE MADRID

Escuela Politécnica Superior



TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

*Sistema avanzado para la indexación y búsqueda de  
recursos audiovisuales basado en metadatos y  
contenidos*

Autor: Carlos Contreras Sanz

Tutor: Jose María Álvarez Rodríguez

Madrid, septiembre 2016



# Índice

Agradecimientos .....	XI
Resumen .....	1
Abstract.....	2
1. Introducción.....	3
1.1. Motivación.....	4
1.2. Objetivos.....	5
1.3. Estructura del documento .....	6
1.4. Acrónimos .....	7
1.5. Definiciones.....	7
2. Estado del arte .....	9
2.1. YouTube .....	9
2.2. Sistemas de reconocimiento de voz.....	10
2.3. Solr.....	11
2.4. Búsquedas contenido multimedia.....	12
3. Análisis del sistema .....	14
3.1. Definición del sistema .....	14
3.2. Análisis de los Casos de Uso .....	15
3.2.1. Roles en el sistema .....	16
3.2.2. Especificación de los casos de uso .....	17
3.3. Definición de requisitos.....	22
3.3.1. Especificación de requisitos Funcionales .....	24
3.3.2. Especificación de requisitos No Funcionales .....	32
3.4. Matriz de trazabilidad.....	36
4. Diseño del sistema .....	38
4.1. Diseño de la Arquitectura .....	38
4.1.1. Modelo.....	39
4.1.2. Vista.....	40
4.1.3. Controlador.....	42
4.2. Diseño de Clases.....	43
4.2.1. Diagrama de Clases .....	43
4.2.2. Identificación de atributos y métodos.....	44
4.2.3. Diagramas de Secuencia.....	49
4.3. Diseño Físico de Datos .....	51
4.3.1. Diseño del modelo físico de la Base de Datos .....	51

4.3.2. Diseño del modelo físico del sistema de ficheros .....	53
4.4. Diseño de interfaces de usuario .....	54
4.4.1. Página principal .....	54
4.4.2. Página de Categorías .....	56
4.4.3. Búsqueda avanzada .....	57
4.5. Especificación del entorno de desarrollo.....	58
4.5.1. Hardware .....	59
4.5.2. Sistemas Operativos .....	59
4.5.3. Lenguajes de programación.....	60
4.5.4. Lenguajes de marcas.....	62
4.5.5. Software de desarrollo .....	64
5. Plan de Pruebas.....	68
5.1. Definición del alcance de las pruebas.....	68
5.2. Entorno de pruebas .....	68
5.3. Especificación de pruebas .....	69
5.3.1. Pruebas unitarias de caja negra.....	69
5.3.2. Pruebas de sistema.....	73
5.4. Análisis de consistencia.....	81
5.5. Estudio de la precisión del sistema.....	82
5.5.1. Precisión y Recall .....	82
6. Implementación .....	86
6.1. Tecnologías utilizadas .....	86
6.1.1. Java Server Pages .....	86
6.1.2. Solr .....	87
6.1.3. Bootstrap.....	90
6.2. Resultado final del sistema .....	91
7. Gestión del proyecto .....	92
7.1. Planificación temporal .....	92
7.1.1. Planificación inicial .....	92
7.1.2. Desarrollo real del Proyecto .....	94
7.2. Presupuesto.....	96
7.2.1. Presupuesto total.....	96
7.2.2. Desglose presupuestario .....	96
7.2.3. Resumen de costes.....	99
8. Marco regulador y ético.....	100

8.1. Java .....	100
8.2. Solr.....	100
8.3. Bootstrap.....	101
8.4. YouTube .....	101
9. Conclusiones y Trabajos Futuros .....	102
9.1. Conclusiones.....	102
9.2. Conclusiones personales.....	103
9.3. Trabajos Futuros .....	104
10. Referencias .....	105
Anexo A: Summary.....	108
Anexo B: Colección de videos .....	116

## Índice de tablas

Tabla 1: Formato de tabla especificación de casos de uso .....	16
Tabla 2: Rol de usuario.....	16
Tabla 3: Rol de administrador .....	17
Tabla 4: Caso de uso CU-01.....	17
Tabla 5: Caso de uso CU-02.....	18
Tabla 6: Caso de uso CU-03.....	19
Tabla 7: Caso de uso CU-04.....	20
Tabla 8: Caso de uso CU-05.....	20
Tabla 9: Caso de uso CU-06.....	21
Tabla 10: Formato tabla descripción de requisitos.....	23
Tabla 11: Requisito RF-01 .....	24
Tabla 12: Requisito RF-02 .....	25
Tabla 13: Requisito RF-03 .....	25
Tabla 14: Requisito RF-04 .....	26
Tabla 15: Requisito RF-05 .....	26
Tabla 16: Requisito RF-06 .....	27
Tabla 17: Requisito RF-07 .....	27
Tabla 18: Requisito RF-08 .....	28
Tabla 19: Requisito RF-09 .....	28
Tabla 20: Requisito RF-10 .....	29
Tabla 21: Requisito RF-11 .....	29
Tabla 22: Requisito RF-12 .....	29
Tabla 23: Requisito RF-13 .....	30
Tabla 24: Requisito RF-14 .....	30
Tabla 25: Requisito RF-15 .....	30
Tabla 26: Requisito RF-16 .....	31
Tabla 27: Requisito RF-17 .....	31
Tabla 28: Requisito RF-18 .....	31
Tabla 29: Requisito RF-22 .....	32
Tabla 30: Requisito RNF-01.....	32
Tabla 31: Requisito RNF-02.....	32
Tabla 32: Requisito RNF-03.....	33
Tabla 33: Requisito RNF-04.....	33
Tabla 34: Requisito RNF-05.....	34

Tabla 35: Requisito RNF-06.....	34
Tabla 36: Requisito RNF-07.....	34
Tabla 37: Requisito RNF-08.....	35
Tabla 38: Requisito RNF-09.....	35
Tabla 39: Matriz de trazabilidad Requisitos/Requisitos.....	36
Tabla 40: Matriz de trazabilidad Requisitos/Casos de Uso .....	37
Tabla 41: Ejemplo tabla definición de Clases .....	44
Tabla 42: Clase CL-01.....	45
Tabla 43: Clase CL-02.....	45
Tabla 44: Clase CL-03.....	45
Tabla 45: Clase CL-04.....	46
Tabla 46: Clase CL-05.....	46
Tabla 47: Clase CL-06.....	47
Tabla 48: Clase CL-07.....	48
Tabla 49: Clase CL-08.....	48
Tabla 50: Configuración ordenador 1.....	59
Tabla 51: Configuración ordenador 2.....	59
Tabla 52: Configuración dispositivo móvil .....	59
Tabla 53: Ejemplo tabla definición prueba unitaria .....	69
Tabla 54: Prueba unitaria PUCN-01.....	70
Tabla 55: Prueba unitaria PUCN-02.....	71
Tabla 56: Prueba unitaria PUCN-03.....	71
Tabla 57: Prueba unitaria PUCN-04.....	72
Tabla 58: Prueba unitaria PUCN-05.....	72
Tabla 59: Ejemplo tabla definición prueba de sistema.....	73
Tabla 60: Prueba de sistema PS-01 .....	74
Tabla 61: Prueba de sistema PS-02 .....	74
Tabla 62: Prueba de sistema PS-03 .....	75
Tabla 63: Prueba de sistema PS-04 .....	75
Tabla 64: Prueba de sistema PS-05 .....	76
Tabla 65: Prueba de sistema PS-06 .....	76
Tabla 66: Prueba de sistema PS-07 .....	77
Tabla 67: Prueba de sistema PS-08 .....	77
Tabla 68: Prueba de sistema PS-09 .....	78
Tabla 69: Prueba de sistema PS-10 .....	78

Tabla 70: Prueba de sistema PS-11 .....	79
Tabla 71: Prueba de sistema PS-12 .....	79
Tabla 72: Prueba de sistema PS-13 .....	80
Tabla 73: Prueba de sistema PS-14 .....	80
Tabla 74: Consistencia Requisitos/Pruebas .....	81
Tabla 75: Resultados consulta 1 .....	83
Tabla 76: Resultados consulta 2 .....	84
Tabla 77: Resultados consulta 3 .....	85
Tabla 78: Costes de personal .....	97
Tabla 79: Costes de material informático.....	98
Tabla 80: Costes de material fungible .....	98
Tabla 81: Costes indirectos .....	98
Tabla 82: Beneficios .....	98
Tabla 83: Resumen de costes.....	99



## Índice de ilustraciones

Ilustración 1: Influencia de las redes sociales por minuto [5] .....	3
Ilustración 2: Visión general del sistema.....	14
Ilustración 3: Caso de uso CU-01 .....	17
Ilustración 4: Caso de uso CU-02.....	18
Ilustración 5: Caso de uso CU-03.....	19
Ilustración 6: Caso de uso CU-06.....	21
Ilustración 7: Modelo-Vista-Controlador [14] .....	39
Ilustración 8: Modelo .....	39
Ilustración 9: Vista .....	40
Ilustración 10: Controlador.....	42
Ilustración 11: Diagrama de Clases .....	43
Ilustración 12: Diagrama de secuencia CU-01 .....	49
Ilustración 13: Diagrama de secuencia CU-02 .....	49
Ilustración 14: Diagrama de secuencia CU-03, CU-04 y CU-05 .....	50
Ilustración 15: Diagrama de secuencia CU-06 .....	50
Ilustración 16: Configuración core SubSearch.....	52
Ilustración 17: Configuración core SubSearchMT .....	52
Ilustración 18: Ejemplo documento core SubSearch.....	53
Ilustración 19: Ejemplo documento core SubSearchMT .....	54
Ilustración 20: Interfaz página principal.....	54
Ilustración 21: Interfaz paginación de resultados .....	55
Ilustración 22: Interfaz estilo de resultados .....	55
Ilustración 23: Desplegable categorías .....	56
Ilustración 24: Interfaz categorías .....	56
Ilustración 25: Formulario búsqueda avanzada .....	57
Ilustración 26: Grafica Precision-Recall consulta 1 .....	83
Ilustración 27: Grafica Precision-Recall consulta 2 .....	84
Ilustración 28: Grafica Precision-Recall consulta 3 .....	85
Ilustración 28: Código ejemplo JSP .....	87
Ilustración 29: Comando acceso al directorio Solr.....	88
Ilustración 30: Comando arranque servidor Solr.....	88
Ilustración 31: Comando creación Core Solr .....	88
Ilustración 32: Comando parar servidor Solr .....	89

Ilustración 33: Comando indexación de documentos.....	89
Ilustración 34: Ejemplo uso clases Bootstrap.....	90
Ilustración 35: Ejemplos visualización dispositivo móvil.....	91
Ilustración 36: Ejemplo visualización pantallas grandes.....	91
Ilustración 37: Planificación inicial estimada.....	93
Ilustración 38: Desarrollo real del proyecto .....	95
Illustration 39: Influence social networks per minute .....	108



## Agradecimientos

En primer lugar, a mi familia por todo el apoyo que me han dado siempre y por estar ahí en todo momento, confiando en mí y ayudándome a seguir. Por su infinita paciencia y comprensión. Sin su ayuda nada de esto hubiera sido posible.

A mis amigos, por aguantarme en los momentos de agobio máximo y saber ofrecer una buena cerveza en los momentos necesarios.

A mis compañeros de carrera y profesores que han hecho que este camino sea más sencillo y fácil de digerir.

Finalmente, a Jose María Alvarez, por sus horas de dedicación como tutor de este proyecto y su actitud predispuesta en todo momento.



## Resumen

El presente trabajo fin de grado tiene como finalidad el crear una página web que permita realizar búsquedas de videos, estos videos tendrán como novedad el uso de la transcripción del video en sus metadatos. Con esto se busca el mejorar los resultados obtenidos al realizar consultas, asemejándose a la forma de búsqueda de páginas web.

Actualmente se desconoce la existencia de un sistema similar que tenga en cuenta las transcripciones a la hora de realizar búsquedas, por lo que en este trabajo se busca plantear la idea y hacer una pequeña demostración de lo que supondría.

Para ello se utilizarán servicios como el de la mayor plataforma de videos online en la actualidad, YouTube. También se utilizarán herramientas con las que se obtendrán los subtítulos de los videos. Además, se desarrollará una herramienta que se encarga de utilizar el servicio Speech to Text de reconocimiento de voz desarrollado por IBM, para demostrar que el reconocimiento de voz automatizado puede ayudar a mejorar las búsquedas de contenido multimedia, en este caso el de los videos.

Por lo tanto, el objetivo de este trabajo es implementar las funcionalidades explicadas, trabajando con una colección de videos limitada, los cuales han sido seleccionados para poder ver a pequeña escala lo que puede llegar a ofrecer esta idea.

## Abstract

This TFG's purpose is to create a webpage that allows users to search videos, these searches will have as novelty the use of the video's transcription contained in their metadata. This seeks improving the results obtained when performing queries, resembling this with webpage's searches.

Nowadays it is not known the existence of a similar system that takes into account the transcriptions when performing searches, so that this work intends to introduce the idea and make a small demonstration of what it would be.

In this project have been used services such as today's largest platform of online videos, YouTube. Tools to obtain the videos' subtitles will also be used. In addition, a tool that uses the voice recognition system Speech to Text, developed by IBM, will be developed to demonstrate that automated voice recognition would help to improve multimedia content searches, videos in this case.

Therefore, the objective of this project is to implement the explained features, working with a limited collection of videos, which have been selected to see in small scale what this idea could offer.

# 1. Introducción

En la actualidad, las tecnologías de la información y la comunicación influyen cada vez más en la sociedad. Esto se debe en gran parte a lo extendido que está el uso del Smartphone en la población, ya que en España un 87% del total de móviles es considerado como teléfono inteligente [1], llegando incluso a superar al número de usuarios que acceden a internet desde ordenadores personales [2].

Además del incremento del número de Smartphone, también se debe a la extensión y facilidad de conexión a internet, ya que en estos momentos existen aproximadamente unos 3500 millones de usuarios de internet en el mundo, entendiendo como usuario a cualquier individuo que tenga acceso a internet desde su casa [3].

Los expertos han llegado a la conclusión, tras estudiar estos hechos, de que los Smartphone son una puerta a las redes sociales [4]. Es por esto por lo que existen compañías que se interesan en estas redes, estudiando el tráfico de las mismas, consiguiendo los siguientes datos [5]:

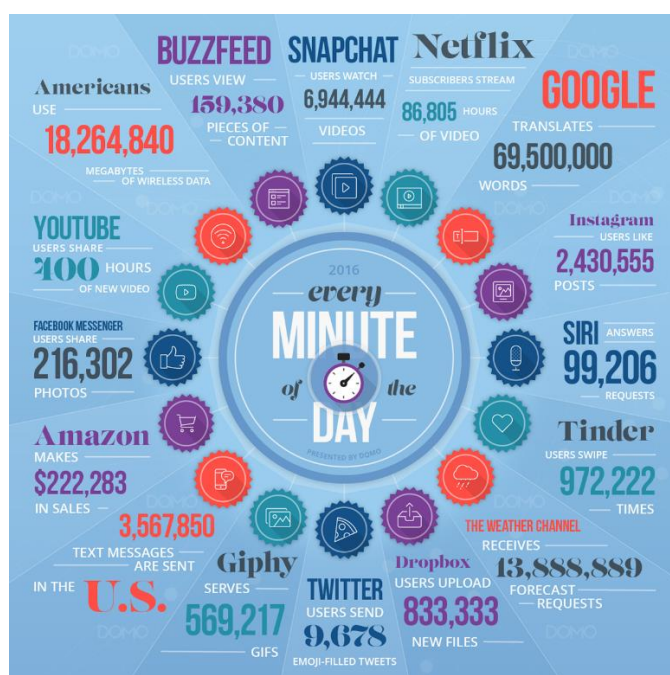


Ilustración 1: Influencia de las redes sociales por minuto [5]

El dato que se pretende dar de forma resumida es el de que cada minuto se generan o consumen una cantidad de información tan grande que hace que surjan necesidades como la necesidad de compresión de archivos, búsqueda de documentos, etc. Es aquí donde se



llega a la necesidad a la que se le pretende proponer una solución con este trabajo de fin de grado, y es la de desarrollar una herramienta que solucione o que mejore las búsquedas sobre contenido multimedia, más concretamente sobre el contenido de videos.

## 1.1. Motivación

Como se puede ver en la introducción, existen servicios como redes sociales, plataformas de video, ... que consumen una cantidad de información por minuto enorme, por ejemplo, los usuarios de YouTube comparten 400 horas de videos, otro ejemplo es el de Netflix, en el que sus usuarios consumen 86.805 horas de video.

Es por esto por lo que estas plataformas, al disponer de tal cantidad de información disponible para sus usuarios buscan ofrecer servicios de búsqueda de calidad, con resultados relevantes. Pero a la hora de la verdad, cuando realizas varias búsquedas el usuario se acaba dando cuenta de que al realizar una búsqueda en los resultados obtenidos aparecen las palabras que ha buscado o bien en el titulo o en la descripción del video, además de que los resultados se ven afectados por otros factores como el número de reproducciones o cantidad de “*Me Gusta*” que reciben.

Es por esto por lo que se piensa en ofrecer un servicio que no tenga tanto en cuenta si aparecen las palabras clave en el titulo o descripción y que, si aparezca en el verdadero contenido, es decir, en el propio video. Esta idea hace unos años igual podría resultar imposible de hacer, ya que no se disponía de los avances tecnológicos actuales, pero en la actualidad existen sistemas de reconocimiento de voz que tienen un margen de error cada vez menor, como es el caso de los subtítulos automáticos de YouTube [6], o herramientas Speech To Text como la de IBM [7].

Además de aplicar esta funcionalidad al mundo de los videos se podría utilizar de igual forma en el mundo de la radio o podcast, ya que se trabajaría con el mismo contenido, ya que el sistema no busca interpretar las imágenes de los videos.

## 1.2. Objetivos

Como se ha venido comentando en los anteriores apartados, el objetivo de este Trabajo Fin de Grado es el diseño de una página web que haga de buscador de videos.

En concreto tiene por objetivo el analizar, diseñar y desarrollar las funciones necesarias para que la página web sea capaz de utilizar un motor de búsqueda que tendrá indexada la información de los videos. Por lo que el sistema se encargara de consultar al motor de búsquedas, recuperar los resultados y presentarlos al usuario.

Además de esto también se busca aumentar la cantidad de videos que ofrezca el sistema, por lo que contara con la función de poder crear la documentación necesaria para la indexación de nuevos videos.

Otro objetivo planteado es el de desarrollar una pequeña aplicación que se encargue de pasar un archivo de audio a texto, utilizando cualquier sistema existente de Speech to Text, para poder utilizarlo a la hora de insertar nuevos videos al índice.

### 1.3. Estructura del documento

En este apartado se hará una breve descripción de cada uno de los capítulos y de la estructura que seguirá el documento.

- **Introducción:** En este capítulo se describen los objetivos y las motivaciones que llevaron a realizarlo, además de enunciar la estructura que seguirá el presente documento.
- **Estado del arte:** En este apartado se analizará el entorno que influye al sistema que se pretende desarrollar. En él se hará un análisis de si existen o no aplicaciones similares, también se analizarán aplicaciones sobre las que se apoya el sistema, sobre las que se explicarán algunos detalles de cada una de ellas... También se comentarán los principales sistemas de reconocimiento de voz o también conocidos como Speech To Text. Y por último se realizará un breve resumen sobre la búsqueda de contenido multimedia en la actualidad.
- **Análisis del sistema:** En este capítulo se realizará un estudio detallado de las funcionalidades a desarrollar en el presente trabajo, en el que se analizarán cada una de las partes con los correspondientes requisitos y casos de uso necesarios para su desarrollo. Se definirá el sistema, y se realizará un contraste entre los requisitos y casos de uso para ver que se han contemplado todas las necesidades. Y por último se realizará una definición de las interfaces de usuario con las que contará el sistema.
- **Diseño del sistema:** En este capítulo se concretará el diseño de la arquitectura, el diseño de las clases y diseño físico de los datos.
- **Plan de pruebas:** En este capítulo se especificará el plan de pruebas, además de realizar algunos análisis del rendimiento del sistema.
- **Implementación:** En este apartado se recogen los aspectos más importantes del desarrollo del proyecto, incluyendo tecnologías utilizadas, esquema del proyecto, etc.
- **Gestión del proyecto:** Este capítulo trata de cómo se planificó el proyecto en un inicio y como ha sido el desarrollo real al final, además de realizar el cálculo del presupuesto del proyecto.
- **Marco regulador y ético:** En este apartado se realizará un pequeño análisis del marco legal que afectaría al sistema en España y Europa.

- **Conclusiones y trabajos futuros:** Por último, se comentarán algunas conclusiones obtenidas tras la realización del presente trabajo, así como las posibles mejoras y líneas futuras del proyecto.

## 1.4. Acrónimos

- **ASF:** *Apache Software Foundation.*
- **CSS:** *Cascading Style Sheets.*
- **ECJ:** *Eclipse Java Compiler.*
- **GPL:** *General Public License.*
- **HTML:** *HyperText Markup Language.*
- **IDE:** *Integrated Development Environment.*
- **JDT:** *Java Development Toolkit.*
- **JRE:** *Java Runtime Environment.*
- **JSP:** *JavaServer Pages.*
- **J2EE:** *Java Platform Enterprise Edition.*
- **J2ME:** *Java Micro Edition.*
- **SGML:** *Standard Generalized Markup Language.*
- **WORA:** *Write Once, Run Anywhere.*
- **W3C:** *World Wide Web Consortium.*
- **XML:** *eXtensible Markup Language.*

## 1.5. Definiciones

- **Apache Lucene:** es una API de código abierto para recuperación de información, originalmente implementada en Java por Doug Cutting. Es útil para cualquier aplicación que requiera indexado y búsqueda a texto completo.
- **Applet:** es un componente de una aplicación que se ejecuta en el contexto de otro programa, por ejemplo, en un navegador web. El applet debe ejecutarse en un contenedor, que le proporciona un programa anfitrión, mediante un plugin, o en aplicaciones como teléfonos móviles que soportan el modelo de programación por "applets".
- **Framework:** es un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular que sirve como referencia, para enfrentar y resolver nuevos problemas de índole similar.

- **Git:** es un software de control de versiones diseñado por Linus Torvalds, pensando en la eficiencia y la confiabilidad del mantenimiento de versiones de aplicaciones cuando éstas tienen un gran número de archivos de código fuente.
- **GNU/Linux:** Es el término empleado para referirse a la combinación del núcleo o kernel libre denominado Linux con el sistema operativo GNU. Todo su código fuente puede ser utilizado, modificado y redistribuido libremente por cualquiera bajo los términos de la Licencia Pública General de GNU y otra serie de licencias libres.
- **Responsive:** El diseño web adaptativo, conocido por las siglas RWD del inglés *Responsive Web Design*, es una filosofía de diseño y desarrollo cuyo objetivo es adaptar la apariencia de las páginas web al dispositivo que se esté utilizando para visualizarlas.
- **Servlet:** es una clase en el lenguaje de programación Java, utilizada para ampliar las capacidades de un servidor, y son utilizados para extender las aplicaciones alojadas en servidores web.
- **Unix:** es un sistema operativo que admite múltiples usuarios, así como también múltiples tareas, lo que significa que permite que en un único equipo o multiprocesador se ejecuten simultáneamente varios programas a cargo de uno o varios usuarios.
- **Webservice:** es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos. La interoperabilidad se consigue mediante la adopción de estándares abiertos.

## 2. Estado del arte

En este apartado se busca situar el entorno que sería más cercano al sistema, analizando las diferentes alternativas posibles que existen para llevar a cabo la idea del sistema que se pretende desarrollar.

### 2.1. YouTube

Al tratarse de una página web que trabaja con videos primero se ha de seleccionar la plataforma de la que se utilizaran los videos, ya que existen varias, como YouTube, Vimeo, Dalymotion... Tras conocer las características de cada una de ellas se llega a la conclusión de forma rápida, ya que YouTube ofrece todo lo necesario para el funcionamiento del sistema que se pretende desarrollar.

YouTube es un sitio web en el cual los usuarios pueden subir y compartir videos. Fue creado por tres antiguos empleados de PayPal en febrero de 2005. En octubre de 2006, fue adquirido por Google Inc. a cambio de 1650 millones de dólares y operando en la actualidad como una de sus filiales.

Actualmente es el sitio web de alojamiento de videos más utilizado en internet, y según el ranking de trafico Alexa es la segunda página web más visitada en todo el mundo [8].

Es por esto por lo que se ha decidido basar el sistema únicamente en los videos de YouTube, ya que es una de las páginas más importantes de internet y además existen aplicaciones externas que pueden ser útiles a la hora de extraer información de los videos.

Además de esto, YouTube cuenta con una API para desarrolladores, por lo que podría utilizarse en el caso de que hiciera falta.

Se puede llegar a pensar que el sistema que se quiere desarrollar pretende hacer lo que YouTube ya hace, pero eso no es así, ya que se conoce la forma en la que YouTube indexa sus videos, y es mediante la de metadatos que el usuario introduce a la hora de subir el video a la plataforma, dándole mayor peso al título y descripción [9].

Es aquí donde el sistema que se pretende desarrollar realiza una innovación a la hora de incluir más información a los metadatos y así poder ofrecer una búsqueda más efectiva o relevante.

Además, se pretende utilizar información de la que el propio YouTube dispone, y son las transcripciones o subtítulos que YouTube almacena para cada video en múltiples idiomas. Teniendo la posibilidad también de conseguir esas transcripciones de manera automática gracias al software de reconocimiento de audio del que disponen.

Por lo que básicamente la idea del sistema sería la de utilizar cierta información obtenida de YouTube como el ID de los videos, título, descripción, fecha y autor. La transcripción de los videos se obtendría con otras herramientas externas de las que se hablara más adelante. Y con toda esta información recopilada crear la documentación necesaria para cada video y dejar que el motor de búsquedas realice su trabajo.

## 2.2. Sistemas de reconocimiento de voz

Otra de las funcionalidades necesarias para el funcionamiento del sistema es el de la existencia de sistemas de reconocimiento de voz, ya que será lo que consiga obtener las transcripciones de los videos. Estas tecnologías actualmente parecen estar de moda, ya que prácticamente cada dispositivo dispone de su asistente por voz, como lo son Siri, Cortana y Google Now.

Existen varios sistemas de reconocimiento de voz diferenciándose entre ellos en los idiomas soportados, porcentaje de error... por lo que tras buscar librerías que pudieran utilizarse, y a su vez ofrecieran el mínimo margen de error posible se llega al servicio Speech to Text del IBM Watson Developer Cloud, un servicio gratuito los mil primeros minutos de cada mes, una vez superados estos minutos pasa a ser de pago, costando \$0.02 el minuto extra [7].

Este servicio soporta gran variedad de idiomas (US Ingles, UK Ingles, Japonés, Castellano, Brasileño Portugués, Modern Standard Árabe, y Mandarín). El uso de este sistema se realiza a través de un Webservice, al que se le envía el .mp3 del que se desea obtener la transcripción y se recibe la transcripción con información adicional como el porcentaje de seguridad de cada palabra, para así poder saber con algo más de certeza si la transcripción automática es buena o es de mala calidad.

Este webservice es el que se utilizara con la herramienta Speech to Text desarrollada para una de las dos posibles vías de obtención de subtítulos.

La segunda vía de obtención de subtítulos es mediante el uso de herramientas externas que se encargan de descargar los subtítulos generados automáticamente por YouTube, siendo este el segundo sistema de reconocimiento de voz que se utilizara en el sistema, en este caso de forma indirecta. El reconocimiento de voz de YouTube ha ido y seguirá mejorando con el tiempo, ya que se trata del mismo software que utiliza Google para su asistente de voz Google Now.

Este apartado, el del reconocimiento automático de voz será una de las claves sobre las que se podrá mejorar búsquedas en el futuro, ya que si se consiguieran márgenes de errores mínimos se podría confiar casi al 100% de la transcripción, por lo que en contenido que se puedan diferenciar claramente las voces como pueden ser programas de radio, monólogos... aportaría información muy valiosa, que si se utiliza de forma adecuada puede ofrecer servicios muy potentes.

## 2.3. Solr

La funcionalidad del sistema es buscar entre los videos de los que dispone, y como tal necesita de un motor de búsqueda para poder realizar búsquedas de forma rápida y sencilla, obteniendo una lista de resultados que encajen con la consulta introducida.

Esto supone escoger entre la oferta de motores de búsqueda que existen, buscando motores de código abierto y a ser posible que tuvieran una API para Java, ya que al tratarse de un proyecto web J2EE se podría desarrollar todo en Java. Es aquí donde aparece Lucene [10], un motor de búsquedas de alto rendimiento desarrollado completamente en Java, el cual ofrece una API que puede ser utilizada por cualquiera. También aparecen los motores de búsqueda Solr [11] y Elasticsearch [12] que utilizando Lucene ofrecen servicios similares. Por lo que, al tratarse de tres motores de búsqueda similares, se seleccionó Solr al parecer que existía más documentación sobre el motor, además de tener una gran comunidad de desarrolladores en redes como StackOverflow y wikis propias en las que se responden numerosas preguntas.

Solr es un motor de búsqueda de código abierto basado en la biblioteca Java del proyecto Lucene, con APIs en XML/HTTP y JSON, resaltado de resultados, búsqueda por facetas, caché, y una interfaz para su administración. Corre sobre un contenedor de servlets Java Apache Tomcat, y ofrece numerosos servicios como:



- Posibilidad de definir el schema de los documentos que serán almacenados.
- Posibilidad de indexar documentos sin seguir un schema definido.
- Permite añadir plugins para indexar nuevos tipos de documentos (PDF, Word, Excel., etc.), detección de idioma...
- Configuración avanzada del análisis de texto, para que el tanto a la hora de indexar, como a la hora de consultar el contenido sea lo más flexible posible.
- Rendimiento optimizado.
- Opciones de almacenamiento avanzado.
- Múltiples índices en el mismo motor de búsqueda.

Algunos de estos servicios han servido de gran utilidad para el sistema, como el de permitir varios índices en el mismo motor de búsqueda, el definir el schema de los documentos a indexar...

## 2.4. Búsquedas contenido multimedia

En definitiva, el trabajo pretende mejorar o al menos plantear una posible mejora de las búsquedas de contenido multimedia, más concretamente el del contenido audiovisual (Videos y Audio únicamente), ya que no se encarga de analizar y/o procesar las imágenes del video.

En la actualidad existen motores de búsqueda como Google, Bing, Yahoo... que ofrecen servicios de búsqueda de páginas web añadiendo también contenido multimedia a sus resultados, como el de imágenes y videos, recuperando estos gracias a funcionalidades ofrecidas por YouTube (videos) y por títulos e información de páginas web en las que aparecen (imágenes). Es decir, en definitiva, siempre se tiende a buscar este tipo de contenido por los típicos metadatos que se utiliza en ellos, como lo son el título, descripción, autor, fecha... sin tener en cuenta el verdadero contenido que es del que se habla durante el video.

Un ejemplo claro es el del podcast, programas similares a programas de radio en los que se habla de diferentes temas. Sobre estos podcasts se puede conocer la información como fecha de grabación, personas que participan, resumen o descripción del programa, título del programa y poco más. Estos son sobre los campos que se consultan, faltando información, la cual se piensa que podría ser clave a la hora de mejorar las búsquedas,

esta información son las transcripciones del audio o video. Gracias a la inclusión de esta meta información se podría dirigir automáticamente a fragmentos del video o audio en los que se habla de lo consultado además de la inclusión de nuevos resultados que podrían pasar desapercibidos al no tener un título y descripción adecuados.

Tras investigarse la existencia de un servicio similar que tuviera en cuenta dichas transcripciones del contenido o características similares, y no encontrar ninguno se considera que este servicio que se pretende ofrecer es en parte innovador al ampliar la información del contenido y así poder enriquecer las búsquedas.

### 3. Análisis del sistema

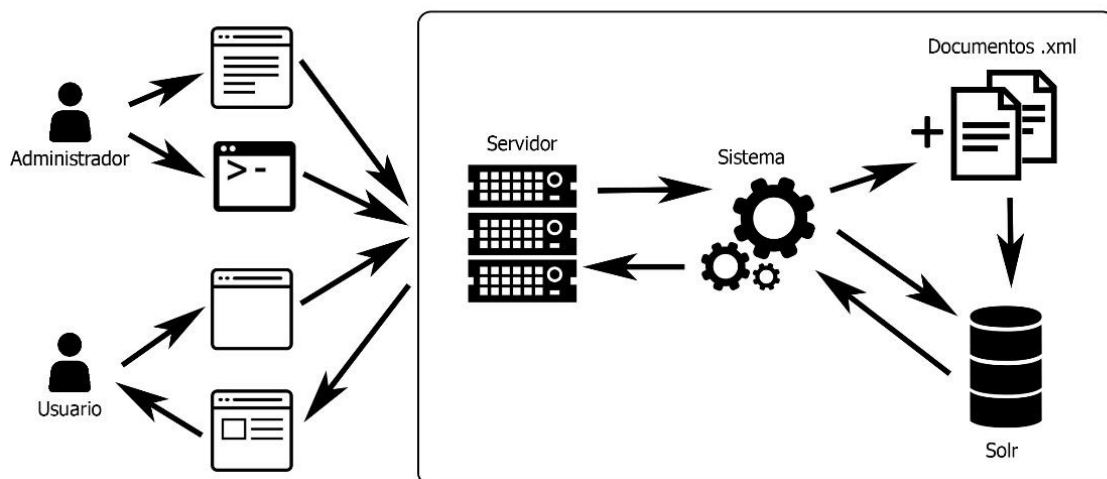
Como se anuncia en Métrica v.3, el Análisis de un sistema de información es “[...] *Obtener una especificación detallada del sistema de información que satisfaga las necesidades de información de los usuarios y sirva de base para el posterior diseño del sistema.*” [13]. Por lo tanto, este apartado va a tratar de conseguir en los siguientes apartados su perfecto análisis.

#### 3.1. Definición del sistema

En este apartado se describirá brevemente el sistema propuesto, delimitar su alcance, así como identificar a los usuarios que toman parte en el sistema.

El sistema que se desea desarrollar consiste en una plataforma web que posee un motor de búsquedas propio, que se encarga de realizar búsquedas sobre los videos que posee. Utilizando como campos de búsqueda su título, descripción, ... dándole más peso a las transcripciones del video, teniendo en cuenta también las marcas temporales.

En la siguiente ilustración se puede apreciar la visión general del sistema:



*Ilustración 2: Visión general del sistema*

Las entidades participantes en la visión general del sistema son las siguientes:

- **Usuario:** Es el usuario que utiliza la página web para realizar una búsqueda.
- **Administrador:** Es el encargado del mantenimiento del sistema, además de tener la posibilidad de añadir videos al sistema.

- **Servidor:** Es el encargado de intercambiar todo tipo de llamadas entre el sistema y los usuarios.
- **Sistema:** Es la pieza fundamental del funcionamiento de la aplicación, ya que será el encargado de diferenciar entre los tipos de llama que le realicen y contestarlas de forma correcta.
- **Documentos “.xml”:** Estos documentos son los que genera el sistema de forma automática cuando recibe la llamada del administrador cuando este desea añadir un video al índice del motor de búsqueda.
- **Solr:** Es el motor de búsqueda utilizado también como base de datos que se encargara de realizar las consultas que le indique el sistema para devolverle los datos que le piden. Además de esto, cuando el administrador decida incluir los videos al índice mediante la consola deberá de obtener los documentos “.xml” generados por la aplicación e indexarlos.

El alcance del sistema completo permitirá a los usuarios realizar consultas desde una página web sobre los videos indexados en el motor de búsquedas y que el sistema se encargue de mostrar los resultados mediante una página web de respuesta. Además, el sistema podrá incluir nuevos videos al índice gracias al administrador, que deberá de seleccionar los videos que se desean de incluir, procesar la información de los mismos e indexarlos de manera manual desde la consola con acceso al servidor.

### 3.2. Análisis de los Casos de Uso

En esta sección se llevará a cabo un análisis exhaustivo de la funcionalidad que se busca ofrecer con el sistema. Con esto se pretende obtener una imagen completa y a alto nivel del sistema captando las necesidades que debe de servir.

Este análisis se consigue con los casos de uso, que consisten en una descripción de las actividades que deberán realizarse para llevar a cabo el uso de la aplicación web por parte los usuarios que deseen utilizarla.

Estos casos de uso se expondrán en una tabla como la siguiente con la que quedara perfectamente definido. Además de esto se incluirá una ilustración del caso de uso para poder ver el orden que se sigue.

<b>Identificador</b>		<b>Nombre</b>	
<b>Actores</b>			
<b>Objetivo</b>			
<b>Escenario</b>			
<b>Condiciones de fallo</b>			

*Tabla 1: Formato de tabla especificación de casos de uso*

Siendo estos sus campos:

- **Identificador:** código identificativo, unívoco, de cada uno de los casos de uso. La nomenclatura utilizada será: **CU-NN**, donde NN es el número que representa el caso de uso.
- **Nombre:** nombre que se le da al caso de uso.
- **Actores:** los posibles roles de usuario que pueden intervenir en el caso de uso.
- **Objetivo:** breve explicación del proceso llevado a cabo en el caso de uso.
- **Escenario:** descripción esquemática de las fases que componen el caso de uso.
- **Condiciones de fallo:** describe posibles errores y las respuestas del sistema ante los mismos.

Antes de especificar los casos de uso se van a enumerar los roles que se considera que tiene el sistema.

### 3.2.1. Roles en el sistema

Debido a las funcionalidades que se pretenden ofrecer con el sistema los roles de usuario que se han considerado son:

Usuario	
<b>Descripción rol</b>	Cada una de las personas que puede utilizar la página web
<b>Funciones</b>	Podrá realizar búsquedas sobre los índices de los que dispone el sistema. Utilizar la búsqueda avanzada, etc.

*Tabla 2: Rol de usuario*

Administrador	
<b>Descripción rol</b>	El administrador del sistema será alguien que conozca el funcionamiento de la misma y esté capacitado para poder solucionar errores que puedan aparecer.
<b>Funciones</b>	Podrá agregar nuevos videos al sistema, utilizar herramientas para conseguir los subtítulos de los videos, obtener el archivo de audio de los videos, etc.

Tabla 3: Rol de administrador

### 3.2.2. Especificación de los casos de uso

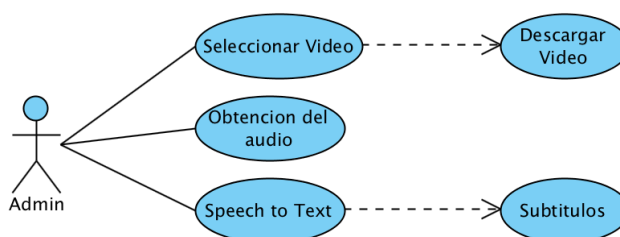


Ilustración 3: Caso de uso CU-01

<b>Identificador</b>	CU-01	<b>Nombre</b>	Speech to Text
<b>Actores</b>	Administrador del sistema		
<b>Objetivo</b>	El administrador del sistema obtiene los subtítulos del video seleccionado utilizando la herramienta Speech to Text desarrollada.		
<b>Escenario</b>	<ol style="list-style-type: none"> <li>1. El Administrador selecciona el video del que desea conseguir los subtítulos.</li> <li>2. El Admin descarga el video.</li> <li>3. El Admin lanza los comandos necesarios en la consola para obtener el archivo de audio del video.</li> <li>4. El Admin lanza los comandos necesarios por el terminal para conseguir el documento que contenga los subtítulos del video.</li> </ol>		
<b>Condiciones de fallo</b>	Error a la hora de conseguir los subtítulos. Se debe de volver a intentar hasta conseguirse, ya que puede ser un error provocado por el servicio utilizado.		

Tabla 4: Caso de uso CU-01

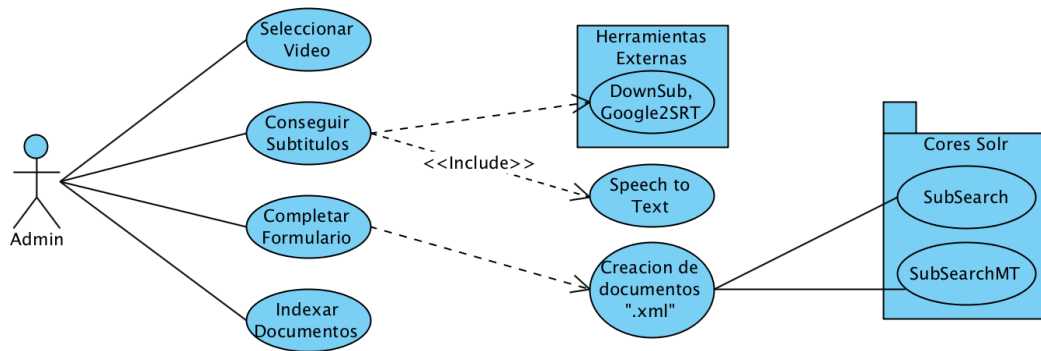


Ilustración 4: Caso de uso CU-02

Identificador	CU-02	Nombre	Añadir video al índice
Actores	Administrador del sistema		
Objetivo	El administrador del sistema añade un nuevo video al índice, aumentando en uno el número de videos del sistema.		
Escenario	<ol style="list-style-type: none"> <li>1. El Administrador selecciona el video que desea añadir a la colección de videos ofrecida por el sistema.</li> <li>2. El Admin consigue los subtítulos del video, ya sea utilizando una herramienta externa o la desarrollada.</li> <li>3. El Admin completa el formulario con los campos ID, Título, Usuario, Nombre de autor, Fecha de publicación, Categoría y Descripción.</li> <li>4. El sistema genera automáticamente los documentos ".xml" necesarios para que el video pueda ser indexado al motor de búsqueda.</li> <li>5. El Admin lanza los comandos necesarios por el terminal para indexar el video al índice utilizado por el sistema.</li> </ol>		
Condiciones de fallo	<ul style="list-style-type: none"> <li>- Error a la hora de conseguir los subtítulos. Se debe de volver a intentar hasta conseguirse.</li> <li>- Falta algún campo del video no conocido. El sistema no permitirá añadir el video.</li> </ul>		

Tabla 5: Caso de uso CU-02

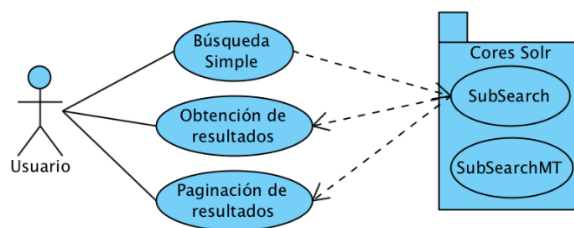


Ilustración 5: Caso de uso CU-03

Identificador	CU-03	Nombre	Búsqueda simple
Actores	Usuario		
Objetivo	El usuario del sistema realiza una consulta sobre la base de datos del motor de búsqueda obteniendo los resultados que mejor se ajustan a la consulta realizada.		
Escenario	<ol style="list-style-type: none"> <li>1. El usuario realiza una consulta en el campo de búsqueda de la barra de navegación.</li> <li>2. El sistema recibe la consulta y la realiza sobre el índice que no tiene en cuenta las marcas de tiempo.</li> <li>3. El sistema obtiene los resultados y los muestra por pantalla, en caso de no encontrar resultados se indicará por pantalla.</li> <li>4. El sistema paginará los resultados en caso de existir más de 10 coincidencias.</li> </ol>		
Condiciones de fallo	<ul style="list-style-type: none"> <li>- Error a la hora de conseguir los resultados. Se lanzarán excepciones en el servidor y al usuario se le mostrara la página de error correspondiente.</li> </ul>		

Tabla 6: Caso de uso CU-03



<b>Identificador</b>	CU-04	<b>Nombre</b>	Selección de Categoría
<b>Actores</b>	Usuario		
<b>Objetivo</b>	El usuario del sistema selecciona la categoría de la que desea ver los resultados, lanzando una consulta sobre la base de datos del motor de búsqueda obteniendo los resultados de la categoría seleccionada.		
<b>Escenario</b>	<ol style="list-style-type: none"> <li>1. El usuario selecciona la categoría que desea de la barra de navegación.</li> <li>2. El sistema recibe la consulta y la realiza sobre el índice que no tiene en cuenta las marcas de tiempo. Ordenando los resultados aleatoriamente para que no aparezcan siempre en el mismo orden.</li> <li>3. El sistema obtiene los resultados y los muestra por pantalla.</li> <li>4. El sistema paginará los resultados en caso de existir más de 10 coincidencias.</li> </ol>		
<b>Condiciones de fallo</b>	Error al seleccionar una categoría de la que no existen resultados. Para ello se conocen con anterioridad las categorías disponibles para evitar este error.		

Tabla 7: Caso de uso CU-04

<b>Identificador</b>	CU-05	<b>Nombre</b>	Acceso a la aplicación web
<b>Actores</b>	Usuario		
<b>Objetivo</b>	El usuario accede a la web, obteniendo en la página principal la lista de videos que existen en el sistema en orden aleatorio para que no aparezcan siempre en el mismo orden.		
<b>Escenario</b>	<ol style="list-style-type: none"> <li>1. El usuario accede a la web.</li> <li>2. El sistema recibe la petición y obtiene la lista de los videos en orden aleatorio.</li> <li>3. El sistema muestra los resultados por pantalla.</li> <li>4. El sistema paginará los resultados en caso de existir más de 10 coincidencias.</li> </ol>		
<b>Condiciones de fallo</b>	Error al intentar recuperar los resultados al no estar arrancado el servidor del motor de búsqueda.		

Tabla 8: Caso de uso CU-05

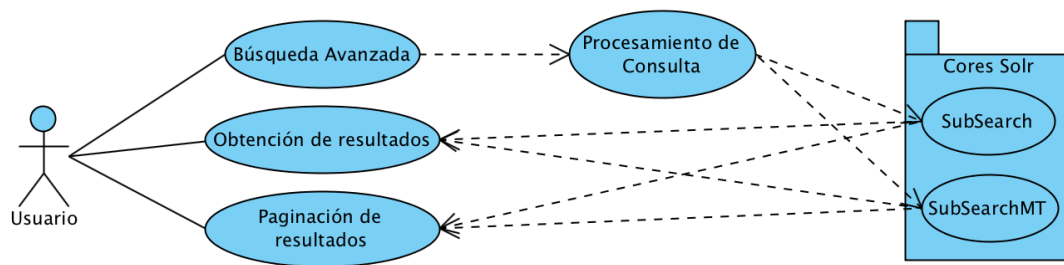


Ilustración 6: Caso de uso CU-06

Identificador	CU-06	Nombre	Búsqueda Avanzada
Actores	Usuario		
Objetivo	El usuario del sistema accede a la Búsqueda Avanzada, y realiza una consulta sobre la base de datos del motor de búsqueda obteniendo los resultados que mejor se ajustan a la consulta realizada.		
Escenario	<ol style="list-style-type: none"> <li>1. El usuario accede a la Búsqueda Avanzada. Y realiza una consulta utilizando los distintos campos de consulta.</li> <li>2. El sistema recoge la información dada por el usuario de cada campo de consulta. Y procesa la consulta a realizar al motor de búsqueda. Esta consulta puede realizarse sobre ambos índices, al existir un parámetro de búsqueda que se puede seleccionar por el usuario, y es el de utilizar marcas de tiempo o no.</li> <li>3. El sistema obtiene los resultados y los muestra por pantalla, en caso de no encontrar resultados se indicará por pantalla.</li> <li>4. El sistema paginará los resultados en caso de existir más de 10 coincidencias.</li> </ol>		
Condiciones de fallo	<ul style="list-style-type: none"> <li>- Error al realizar consultas que puedan afectar al funcionamiento de Solr. Para solucionarlo se han escapado caracteres considerados como especiales para evitar este error.</li> <li>- Error a la hora de no poder completar el formulario al utilizar navegadores web no soportados por el sistema.</li> </ul>		

Tabla 9: Caso de uso CU-06

### 3.3. Definición de requisitos

En este apartado se expondrá el catálogo de requisitos del sistema. Estos requisitos se dividen en dos grandes grupos:

- **Requisitos Funcionales:** Este tipo de requisitos definen las funciones del sistema software o sus componentes. Estos pueden ser: Cálculos, detalles técnicos, manipulación de datos, entre otras funcionalidades que el sistema debe de cumplir.
- **Requisitos No Funcionales:** Los requisitos no funcionales son aquellos requisitos que especifica criterios que pueden usarse para juzgar la operación de un sistema. Es decir, son todos los requisitos que no describen características de funcionamiento, como pueden ser el rendimiento, accesibilidad, etc.

Estos requisitos a su vez podrán pertenecer a diferentes categorías dependiendo del enfoque del requisito. Estas son las categorías que podrán tener los requisitos:

- Requisitos Funcionales:
  - Sistema
  - Interfaz
  - Idioma
- Requisitos No Funcionales:
  - Interfaz
  - Recursos
  - Operatividad

Cada uno de los requisitos se describirán en forma tabular siguiendo el formato de la siguiente tabla. Se ha decidido seguir esta estructura de ordenación para mostrar todos los requisitos de forma clara y concisa de forma que posteriormente puedan testearse.

Identificador		Categoría	
Nombre			
Prioridad	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
Riesgo	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
Verificabilidad	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
Descripción			
Requisitos relacionados			

Tabla 10: Formato tabla descripción de requisitos

A continuación, se describen los campos de la tabla:

- **Identificador:** es el código identificativo, unívoco, de cada uno de los requisitos. La nomenclatura utilizada será: **RX-NN**. El significado de cada una de las letras es el siguiente:
  - **R:** indica que es un requisito
  - **X:** indica el tipo de requisito a elegir entre los siguientes tipos:
    - **F:** Requisito Funcional: describe la operación que debe realizar la aplicación.
    - **NF:** Requisito No Funcional: indica cómo debe realizarse la operación.
  - **NN:** es el número de requisito dentro de la clasificación.
- **Categoría:** categoría correspondiente al requisito.
- **Nombre:** Este campo deberá de expresar con un breve título un resumen de la descripción del requisito.
- **Prioridad:** indica el nivel de atención al requisito y la necesidad de satisfacción del mismo. Hay tres niveles:
  - **Alta:** el requisito debe ser satisfecho antes que otros requisitos de prioridad más baja.
  - **Media:** el requisito con este nivel debe ser satisfecho después de haber cumplido con los de nivel alto.
  - **Baja:** requisito que, en virtud de su nivel de prioridad, es el último en satisfacerse.

- **Riesgo:** indica la relación del requisito con su dificultad y necesidad dentro del sistema. Tiene también tres niveles, como los indicados en el punto anterior.
- **Verificabilidad:** indica la facilidad de comprobación que se cumple el requisito. Tiene tres niveles:
  - **Alta:** el requisito puede comprobarse mediante todas las herramientas de verificación.
  - **Media:** el requisito no puede comprobarse mediante todas las herramientas de verificación aplicables.
  - **Baja:** requisito que no puede verificarse mediante las herramientas disponibles.
- **Descripción:** descripción completa del requisito indicando todos sus aspectos y características.
- **Requisitos relacionados:** indica el identificador de los requisitos con los que está si los hay, sino quedará en blanco este campo

### 3.3.1. Especificación de requisitos Funcionales

<b>Identificador</b>	RF-01	<b>Categoría</b>	Sistema
<b>Nombre</b>	Unidad de trabajo		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema trabajara con videos obtenidos de YouTube.		
<b>Requisitos relacionados</b>			

Tabla 11: Requisito RF-01

<b>Identificador</b>	RF-02	<b>Categoría</b>	Sistema
<b>Nombre</b>	Motor de búsqueda		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema utilizara el motor de búsquedas Apache Solr en su versión 6.0.0.		
<b>Requisitos relacionados</b>			

Tabla 12: Requisito RF-02

<b>Identificador</b>	RF-03	<b>Categoría</b>	Sistema
<b>Nombre</b>	Campos de los videos		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	<p>Los videos deberán de poseer los siguientes campos:</p> <ul style="list-style-type: none"> <li>- <u>ID</u>: identificador único que tiene el video en YouTube.</li> <li>- <u>Título</u>: título del video.</li> <li>- <u>Autor</u>: nombre del autor del video.</li> <li>- <u>Usuario</u>: usuario del autor del video.</li> <li>- <u>Fecha de publicación</u>: fecha en la que el video fue publicado.</li> <li>- <u>Categoría</u>: categoría del video asignada por YouTube.</li> <li>- <u>Descripción</u>: descripción del video.</li> <li>- <u>Contenido</u>: transcripciones del audio del video.</li> <li>- <u>Marca de tiempo</u>: marca de tiempo en la que se producen los comentarios en el video.</li> </ul>		
<b>Requisitos relacionados</b>	RF-01		

Tabla 13: Requisito RF-03

<b>Identificador</b>	RF-04	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsqueda de videos		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema será capaz de realizar búsquedas sobre todos los campos de los videos que dispone.		
<b>Requisitos relacionados</b>	RF-02, RF-03		

Tabla 14: Requisito RF-04

<b>Identificador</b>	RF-05	<b>Categoría</b>	Sistema
<b>Nombre</b>	Secciones del sistema		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	<p>El sistema dispondrá de una interfaz con las siguientes secciones o apartados dentro de la misma:</p> <ul style="list-style-type: none"> <li>- Búsqueda simple</li> <li>- Búsqueda avanzada</li> <li>- Sección de Categorías</li> </ul>		
<b>Requisitos relacionados</b>			

Tabla 15: Requisito RF-05

<b>Identificador</b>	RF-06	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsquedas simples		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema permitirá realizar búsquedas simples desde cualquier sección del sistema. Estas búsquedas no tendrán en cuenta el campo marca de tiempo.		
<b>Requisitos relacionados</b>	RF-04, RF-05		

Tabla 16: Requisito RF-06

<b>Identificador</b>	RF-07	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsquedas avanzadas		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	<p>El sistema permitirá realizar búsquedas avanzadas mediante un formulario que tendrá los siguientes apartados que podrán incluirse y combinarse según decida el usuario:</p> <ul style="list-style-type: none"> <li>- Búsqueda por palabras</li> <li>- Búsqueda por marcas de tiempo</li> <li>- Búsqueda por categorías</li> <li>- Búsqueda por autor</li> <li>- Búsqueda por fecha</li> </ul>		
<b>Requisitos relacionados</b>	RF-04, RF-05		

Tabla 17: Requisito RF-07



<b>Identificador</b>	RF-08	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsqueda avanzada por palabras		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	<p>La búsqueda avanzada por palabras incluirá los siguientes campos de entrada:</p> <ul style="list-style-type: none"> <li>- Todas estas palabras: palabras que se desea que aparezcan en los resultados.</li> <li>- Cualquiera de estas palabras: palabras que pueden aparecer o no en los resultados.</li> <li>- Ninguna de estas palabras: palabras que se desean que no aparezcan en los resultados.</li> </ul>		
<b>Requisitos relacionados</b>	RF-07		

Tabla 18: Requisito RF-08

<b>Identificador</b>	RF-09	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsqueda avanzada por marcas de tiempo		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	<p>La búsqueda avanzada por marcas de tiempo permitirá al usuario elegir entre que los resultados le lleven al momento del video (Con Marcas de Tiempo) en el que encaja la búsqueda o que únicamente le dé como resultado el video (Sin Marcas de Tiempo). Los resultados con marcas de tiempo pueden contener varias veces el mismo video.</p>		
<b>Requisitos relacionados</b>	RF-07		

Tabla 19: Requisito RF-09

<b>Identificador</b>	RF-10	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsqueda avanzada por categoría		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	La búsqueda avanzada por categoría permitirá al usuario elegir mediante una lista la categoría de la que desea obtener los resultados. Existe la posibilidad de no seleccionar ninguna y así poder consultar en todas las categorías.		
<b>Requisitos relacionados</b>	RF-07		

Tabla 20: Requisito RF-10

<b>Identificador</b>	RF-11	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsqueda avanzada por autor		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	La búsqueda avanzada por autor permitirá al usuario buscar videos que hayan sido creados por el autor que introduzca en el campo. Esto incluye a los campos del video de usuario y autor.		
<b>Requisitos relacionados</b>	RF-07		

Tabla 21: Requisito RF-11

<b>Identificador</b>	RF-12	<b>Categoría</b>	Sistema
<b>Nombre</b>	Búsqueda avanzada por fecha		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	La búsqueda avanzada por fecha permitirá al usuario buscar videos que hayan sido subidos a YouTube antes de una fecha, después de una fecha y entre dos fechas.		
<b>Requisitos relacionados</b>	RF-07		

Tabla 22: Requisito RF-12

<b>Identificador</b>	RF-13	<b>Categoría</b>	Sistema
<b>Nombre</b>	Enlace a video de YouTube		
<b>Prioridad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema enlazará los videos disponibles en la plataforma con los videos de YouTube abriéndolos en una pestaña nueva.		
<b>Requisitos relacionados</b>	RF-01		

Tabla 23: Requisito RF-13

<b>Identificador</b>	RF-14	<b>Categoría</b>	Sistema
<b>Nombre</b>	Enlace a canal del autor del video		
<b>Prioridad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema enlazará los videos disponibles en la plataforma con los canales de los usuarios que han creado y subido el video en YouTube, abriéndolos en una pestaña nueva.		
<b>Requisitos relacionados</b>	RF-01		

Tabla 24: Requisito RF-14

<b>Identificador</b>	RF-15	<b>Categoría</b>	Interfaz
<b>Nombre</b>	Barra de navegación		
<b>Prioridad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema dispondrá de una barra de navegación con accesos a las distintas secciones del sistema, que será idéntica en cada una de las vistas del mismo.		
<b>Requisitos relacionados</b>	RF-05		

Tabla 25: Requisito RF-15

<b>Identificador</b>	RF-16	<b>Categoría</b>	Idioma
<b>Nombre</b>	Idiomas de videos soportados		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema permitirá únicamente videos cuyos idiomas sean el inglés y castellano.		
<b>Requisitos relacionados</b>	RF-01		

Tabla 26: Requisito RF-16

<b>Identificador</b>	RF-17	<b>Categoría</b>	Sistema
<b>Nombre</b>	Formulario para añadir video		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema dispondrá de un formulario con los campos de un video, permitiendo añadir videos al sistema.		
<b>Requisitos relacionados</b>	RF-03		

Tabla 27: Requisito RF-17

<b>Identificador</b>	RF-18	<b>Categoría</b>	Sistema
<b>Nombre</b>	Generación archivos de indexación		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Verificabilidad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema generara de manera automática los documentos ".xml" en el directorio indicado, utilizando la información del formulario, para disponer de una copia de seguridad y poder indexar el video.		
<b>Requisitos relacionados</b>	RF-17		

Tabla 28: Requisito RF-18

<b>Identificador</b>	RF-19	<b>Categoría</b>	Sistema
<b>Nombre</b>	Limpieza de transcripciones		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema será capaz de limpiar las transcripciones de caracteres no deseados.		
<b>Requisitos relacionados</b>	RF-17		

Tabla 29: Requisito RF-22

### 3.3.2. Especificación de requisitos No Funcionales

<b>Identificador</b>	RNF-01	<b>Categoría</b>	Interfaz
<b>Nombre</b>	Framework de desarrollo		
<b>Prioridad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	La interfaz deberá de realizarse utilizando el framework de desarrollo responsive que ofrece Bootstrap.		
<b>Requisitos relacionados</b>			

Tabla 30: Requisito RNF-01

<b>Identificador</b>	RNF-02	<b>Categoría</b>	Interfaz
<b>Nombre</b>	Idioma interfaz		
<b>Prioridad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	La interfaz del sistema tendrá como único idioma el español o castellano.		
<b>Requisitos relacionados</b>			

Tabla 31: Requisito RNF-02

<b>Identificador</b>	RNF-03	<b>Categoría</b>	Interfaz
<b>Nombre</b>	Lista de resultados		
<b>Prioridad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema mostrara los resultados de las consultas en forma de lista. Siendo 10 el número máximo de videos por vista.		
<b>Requisitos relacionados</b>	RF-04		

Tabla 32: Requisito RNF-03

<b>Identificador</b>	RNF-04	<b>Categoría</b>	Interfaz
<b>Nombre</b>	Elementos del video en la interfaz		
<b>Prioridad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema mostrara por pantalla los siguientes elementos de un video: <ul style="list-style-type: none"> <li>- Miniatura del video (Incluyendo enlace al video)</li> <li>- Título del video (Incluyendo enlace al video)</li> <li>- Autor del video (Incluyendo enlace al canal)</li> <li>- Descripción del video</li> </ul>		
<b>Requisitos relacionados</b>	RNF-03, RF-13, RF-14		

Tabla 33: Requisito RNF-04

Identificador	RNF-05	Categoría	Interfaz
Nombre	Paginación de videos		
Prioridad	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
Riesgo	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
Verificabilidad	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
Descripción	El sistema realizara una paginación de los resultados en caso de superar los 10 resultados.		
Requisitos relacionados	RNF-03		

Tabla 34: Requisito RNF-05

Identificador	RNF-06	Categoría	Interfaz
Nombre	Búsqueda avanzada mínima		
Prioridad	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
Riesgo	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
Verificabilidad	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
Descripción	La búsqueda avanzada no se podrá lanzar si no se han introducido los suficientes parámetros de entrada como para realizar una consulta.		
Requisitos relacionados	RF-07		

Tabla 35: Requisito RNF-06

Identificador	RNF-07	Categoría	Operatividad
Nombre	Usuario administrador		
Prioridad	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
Riesgo	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
Verificabilidad	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
Descripción	El sistema necesitara un usuario administrador que sea el encargado de añadir videos a la plataforma.		
Requisitos relacionados			

Tabla 36: Requisito RNF-07

<b>Identificador</b>	RNF-08	<b>Categoría</b>	Recursos
<b>Nombre</b>	Obtención de transcripciones (I)		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Verificabilidad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema podrá obtener las transcripciones con marcas de tiempo del video que se desea utilizando la aplicación Speech to Text desarrollada.		
<b>Requisitos relacionados</b>	RF-17		

Tabla 37: Requisito RNF-08

<b>Identificador</b>	RNF-09	<b>Categoría</b>	Recursos
<b>Nombre</b>	Obtención de transcripciones (II)		
<b>Prioridad</b>	<input checked="" type="checkbox"/> Alta	<input type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Riesgo</b>	<input type="checkbox"/> Alta	<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Baja
<b>Verificabilidad</b>	<input type="checkbox"/> Alta	<input checked="" type="checkbox"/> Media	<input type="checkbox"/> Baja
<b>Descripción</b>	El sistema podrá obtener las transcripciones de los videos utilizando aplicaciones externas que permiten descargar los subtítulos de los videos.		
<b>Requisitos relacionados</b>	RF-17		

Tabla 38: Requisito RNF-09



### 3.4. Matriz de trazabilidad

En este apartado se realizará un estudio para detectar inconsistencias, ambigüedad, duplicidad o escasez de información. Dicho estudio consistirá en el análisis de los requisitos y de los casos de uso, asociándolos para identificar sus funcionalidades y comportamientos comunes.

Para ello se realizarán dos matrices de trazabilidad, la primera será entre los requisitos, para ver las relaciones entre ellos, y la segunda entre los distintos casos de uso y los requisitos. De esta manera se asegurará de que los requisitos especificados en el catálogo de requisitos son válidos, consistentes y completos, y de que al menos existe un requisito por cada caso de uso definido anteriormente.

A continuación, se muestra la matriz de trazabilidad que muestra las relaciones entre los Requisitos funcionales y no funcionales:

	RF-01	RF-02	RF-03	RF-04	RF-05	RF-06	RF-07	RF-08	RF-09	RF-10	RF-11	RF-12	RF-13	RF-14	RF-15	RF-16	RF-17	RF-18	RF-19	RNF-01	RNF-02	RNF-03	RNF-04	RNF-05	RNF-06	RNF-07	RNF-08	RNF-09
RF-01			X										X	X		X												
RF-02				X																								
RF-03	X			X													X											
RF-04		X	X			X	X														X							
RF-05						X	X								X													
RF-06				X	X																							
RF-07				X	X			X	X	X	X	X													X			
RF-08							X																					
RF-09							X																					
RF-10							X																					
RF-11							X																					
RF-12							X																					
RF-13	X																						X					
RF-14	X																						X					
RF-15					X																							
RF-16	X																											
RF-17			X															X	X								X	X
RF-18																		X										
RF-19																		X										
RNF-01																												
RNF-02																												
RNF-03				X																			X	X				
RNF-04													X	X								X						
RNF-05																						X						
RNF-06							X																					
RNF-07																												
RNF-08																		X										
RNF-09																		X										

Tabla 39: Matriz de trazabilidad Requisitos/Requisitos

A continuación, se muestra la matriz de trazabilidad de los Requisitos funcionales y los Casos de Uso:

	CU-01	CU-02	CU-03	CU-04	CU-05	CU-06
RF-01	X	X	X	X	X	X
RF-02		X	X	X	X	X
RF-03		X	X	X		X
RF-04			X	X	X	X
RF-05			X	X		X
RF-06			X			
RF-07						X
RF-08						X
RF-09						X
RF-10						X
RF-11						X
RF-12						X
RF-13			X	X	X	X
RF-14			X	X	X	X
RF-15			X	X	X	X
RF-16		X				
RF-17		X				
RF-18		X				
RF-19	X	X				
RNF-01		X	X	X	X	X
RNF-02		X	X	X	X	X
RNF-03			X	X	X	X
RNF-04			X	X	X	
RNF-05			X	X	X	
RNF-06						X
RNF-07	X	X				
RNF-08	X	X				
RNF-09	X	X				

Tabla 40: Matriz de trazabilidad Requisitos/Casos de Uso

Como se puede apreciar con un simple vistazo, los requisitos completan la funcionalidad que se busca ofrecer según los casos de uso, ya que todos los casos de uso tienen al menos un requisito asociado, al igual que cada uno de los requisitos tiene asociado algún caso de uso.

## 4. Diseño del sistema

A lo largo de este capítulo se realizará un diseño en el que se definirá la arquitectura del sistema. Por otro lado, se definirá el entorno tecnológico que le dará forma a dicho sistema, así como especificar de forma más detallada los distintos componentes del sistema.

Primeramente, se explicarán los detalles de cada uno de los componentes que conforman el sistema, así como sus funciones, el diseño de clases, el diseño físico de datos, las interfaces de usuario, la forma que tendrán los ficheros y las bases de datos.

### 4.1. Diseño de la Arquitectura

En este apartado se determinarán los distintos módulos que constituirán los subsistemas de soporte que serán utilizados por el sistema desarrollado. El diseño del sistema se corresponderá con el patrón de diseño Modelo-Vista-Controlador, ya que se define una organización independiente del Modelo (Objetos de Negocio), la Vista (Interfaz con el usuario u otro sistema) y el Controlador (controlador del Workflow de la aplicación) [14]. Quedando clasificadas de la siguiente manera:

- **MODELO:** Estará formado por el subsistema de datos, conteniendo el núcleo de la funcionalidad del sistema. Esto quiere decir que será el encargado de proporcionar al resto de capas la información necesaria para consultar o manipular la información con la que trabaja esta capa.
- **VISTA:** Estará formada por el subsistema de interfaces, ya que su función es la de presentar la información del Modelo al usuario. Además de presentar información también es capaz de interactuar con los usuarios, al poder recoger inputs de los mismos.
- **CONTROLADOR:** Estará formado por el subsistema de la lógica operacional, ya que será el encargado de reaccionar a la petición del cliente, teniendo que ejecutar la acción adecuada y creando el modelo pertinente.

Este patrón de arquitectura ofrece una gran transparencia al hacer que cada una de las tres capas sean independientes del resto, por lo que además de la transparencia aporta también ventajas a la hora de que se produzcan fallos en el sistema, permitiendo encontrar el fallo en una de las tres capas, quedando delimitado.

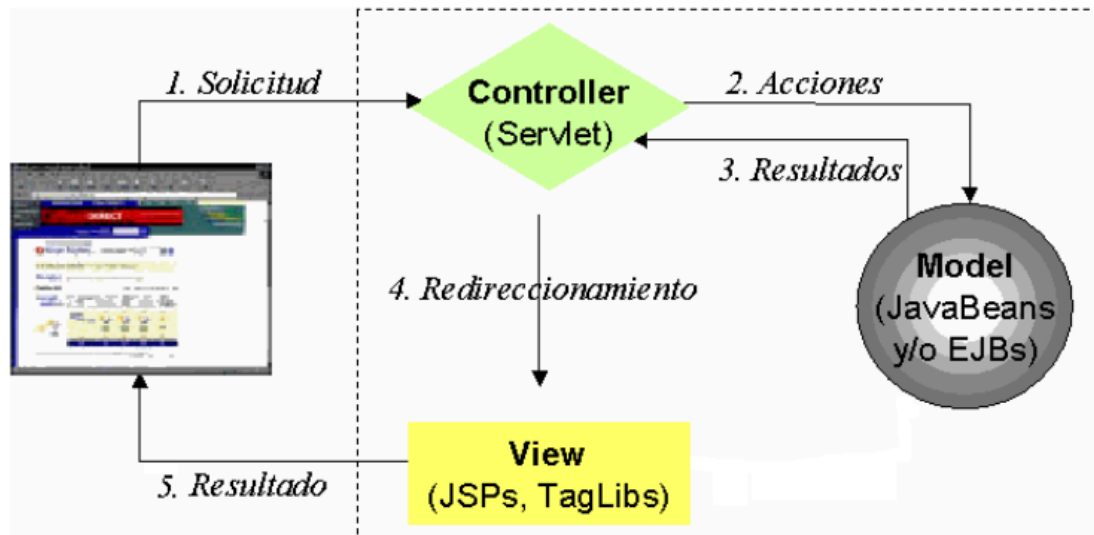


Ilustración 7: Modelo-Vista-Controlador [14]

#### 4.1.1. Modelo

Como se ha comentado en el apartado anterior esta capa se diferencia del resto en que contiene el núcleo de la funcionalidad, por lo que se podría decir que encapsula el estado del sistema.

Este componente es el encargado de realizar las operaciones responsables de la gestión de almacenamiento y consulta de datos sobre la base de datos, inserción de nuevos datos, modificarlos y borrarlos.

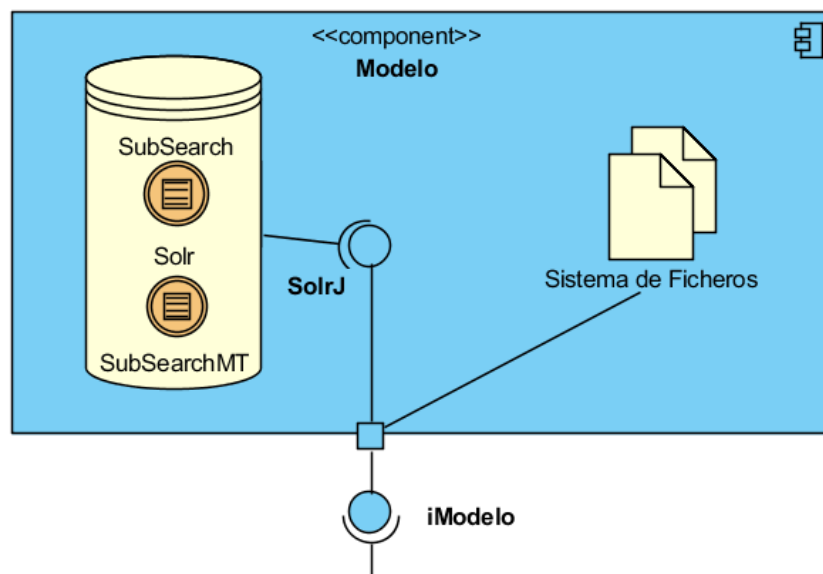


Ilustración 8: Modelo

- **Solr:** es el motor de búsquedas escogido, por lo que será el encargado de indexar y almacenar toda la información de los videos que contenga el sistema. Destacar que Solr permite tener varios índices a la vez, por lo que el sistema al ofrecer una búsqueda teniendo en cuenta las marcas de tiempo y otra sin este parámetro, ha sido necesario el uso de dos índices como se puede ver en la ilustración. La forma de conexión entre el controlador y Solr es mediante el uso de la API SolrJ.
- **Sistema de Ficheros:** es el “almacén” donde se guardarán todos los ficheros necesarios creados automáticamente por el sistema para la indexación de los videos sobre los índices de Solr.

#### 4.1.2. Vista

La capa de Vista es la encargada de presentar el Modelo de una forma más amigable para los usuarios del sistema, permitiéndoles la interacción entre el usuario y el sistema, es decir, se trata de la interfaz gráfica.

En este caso al tratarse de un proyecto que busca crear una aplicación web cuyo uso va a estar basado en sistemas de servidor, se utilizarán documentos JSP para la creación de estas vistas que se generaran de forma dinámica.

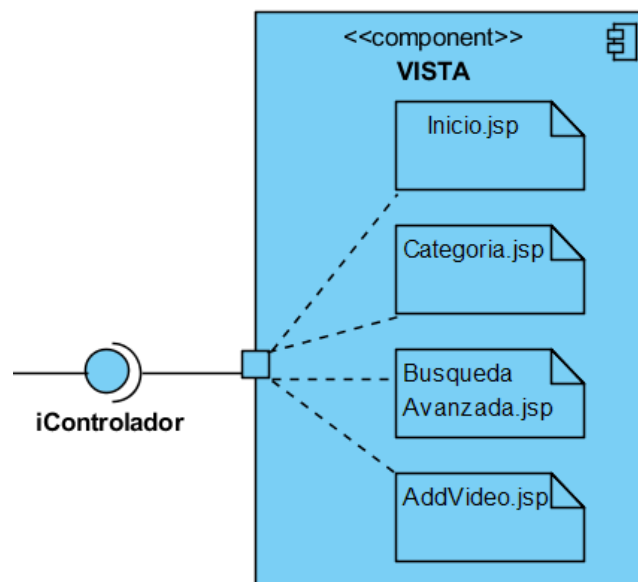


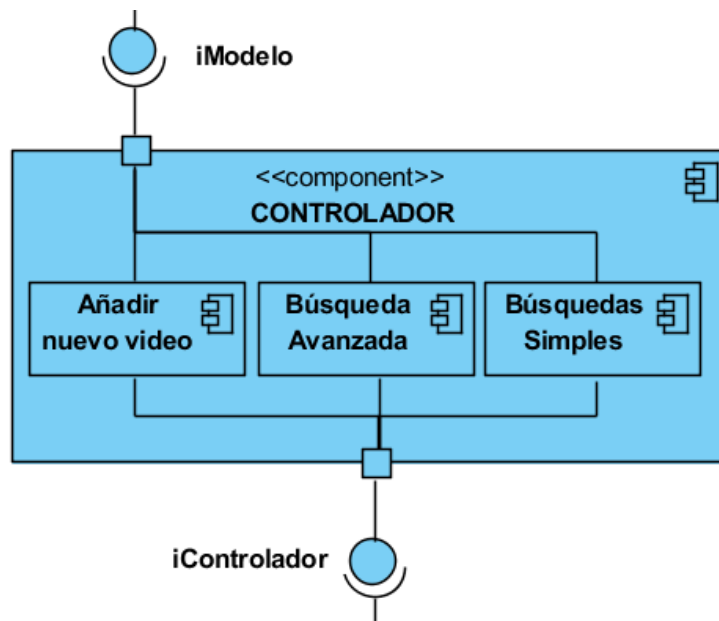
Ilustración 9: Vista

Como se puede apreciar en la ilustración el componente utiliza la información que le llega del Controlador para así mostrar al usuario la interfaz que le corresponde. Las interfaces JSP con las que cuenta el sistema son las siguientes:

- **Inicio.jsp:** Se trata de la página de inicio del sistema, pagina que será utilizada también para mostrar los resultados de las búsquedas avanzadas y búsquedas simples.
- **Categoria.jsp:** Se trata de la página en la que se mostraran los videos de cada una de las categorías disponibles.
- **BúsquedaAvanzada.jsp:** Se trata de la página desde la que el usuario podrá introducir los datos de la consulta que desea realizar, recogiendo la información y dándosela al Controlador para que este sepa qué hacer con ella.
- **AddVideo.jsp:** Se trata de la vista desde la que el usuario administrador tendrá acceso para rellenar los campos del formulario y proceder a añadir un nuevo video al índice.

### 4.1.3. Controlador

En último lugar se encuentra la capa del Controlador, siendo la encargada de recibir y responder cada uno de los eventos que se produzcan en el sistema. Es decir, es la encargada de procesar las entradas que el usuario realiza, consultarlas con la capa de Modelo y finalmente pedirle a la vista que muestre los datos obtenidos de la capa de Modelo, siendo el componente que se encarga de la lógica de operación del sistema.



*Ilustración 10: Controlador*

En esta ilustración se pueden apreciar los siguientes componentes incluidos dentro del Controlador:

- **Añadir nuevo video:** Este componente es el encargado de recoger la información recibida del formulario para añadir un nuevo video a la colección. Una vez recogida la procesa realizando varias operaciones sobre algunos campos, como el de los subtítulos, y una vez preparados se encarga de darle al Modelo la orden de crear los documentos “.xml” necesarios para añadir el video.
- **Búsqueda avanzada:** Este componente es el encargado de recoger los datos del formulario de búsqueda avanzada que introduzca el usuario. Una vez tenga los datos los procesara y se los pasara a la capa de Modelo, esperando a que esta le responda. Cuando recoja la respuesta enviara la información resultante a la capa de Vista para que el usuario vea los resultados de la consulta.

- **Búsquedas Simples:** Básicamente su funcionamiento es idéntico al de la búsqueda avanzada el único cambio significativo es el del número de entradas que recibe, que al ser menor apenas necesita procesamiento.

## 4.2. Diseño de Clases

Este apartado tiene como objetivo generar el modelo de clases de diseño del sistema. El modelo recogerá detalladamente las especificaciones de cada una de las clases basándose en los escenarios contemplados en los casos de uso y arquitectura anunciada en el apartado anterior.

### 4.2.1. Diagrama de Clases

Al basarse la arquitectura del sistema en una arquitectura que sigue el patrón Modelo-Vista-Controlador, se distinguirán las clases que pertenezcan a cada una de las capas, quedando claramente separadas en el diagrama de clases del sistema.

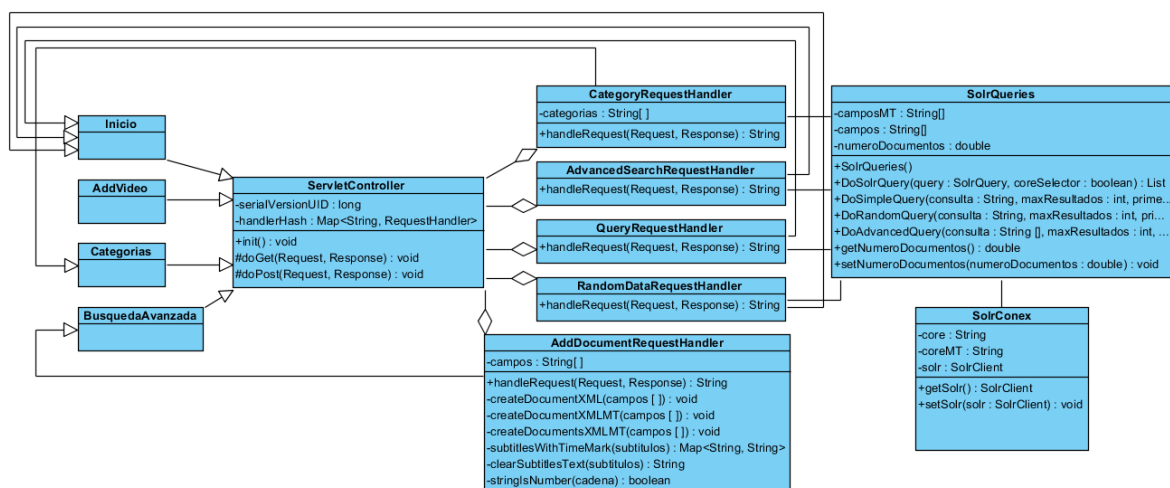


Ilustración 11: Diagrama de Clases

Se puede apreciar de un simple vistazo que la capa de Vista no incluye métodos ni atributos, ya que únicamente se limita a presentar la información que le comunican las clases Handler. Las clases centrales son las que forman la parte del Controlador.

Como se puede ver en la imagen, las cuatro clases de la Vista se unen con la clase ServletController, ya que esta es la encargada de dirigir las llamadas entrantes de los usuarios hacia las clases que sepan responder a cada una de las llamadas, es por esto por



lo que existen las clases `RequestHandler`, que son las que reciben la información y la procesan para comunicarse con la capa de Modelo.

Esta última capa está formada por las clases `SolrQueries` y `SolrConex`, ambas encargadas de realizar cambios sobre la base de datos Solr.

Una vez la capa de Modelo responda a la clase `RequestHandler` correspondiente, esta se encargará de mostrar al usuario la vista correspondiente a la información solicitada.

#### 4.2.2. Identificación de atributos y métodos

A pesar de estar representados en el diagrama de clases anterior, en este apartado se identificarán los métodos y atributos de cada una de las clases que se representa en el diagrama de clases, a excepción de las clases que forman la capa de vista, ya que no poseen nada, como se explica en el enunciado anterior.

Esta identificación se realizará mediante el uso de las siguientes tablas:

CL-NN	
Nombre de clase	
Descripción	
Atributos	
Métodos	

*Tabla 41: Ejemplo tabla definición de Clases*

Siendo sus campos los siguientes:

- **CL-NN:** código identificativo, unívoco, de cada uno de las clases. La nomenclatura utilizada será: **CL-NN**, donde NN es el número que representa a la clase.
- **Nombre de clase:** nombre de la clase a definir.
- **Descripción:** resumen de la finalidad de la clase.
- **Atributos:** lista de atributos que contiene la clase.
- **Métodos:** lista de métodos que contiene la clase.

#### 4.2.2.1 Clases que forman la capa de Controlador

CL-01	
<b>Nombre de clase</b>	ServletController
<b>Descripción</b>	Esta clase es el Servlet que se encarga de recibir todas las peticiones que se realizan desde cualquiera de las páginas del sistema, tras recibir la petición se la pasa a la clase que le corresponde responderla.
<b>Atributos</b>	<ul style="list-style-type: none"><li>- long serialVersionUID</li><li>- Map&lt;String, RequestHandler&gt; handlerHash</li></ul>
<b>Métodos</b>	<ul style="list-style-type: none"><li>- void init ()</li><li>- void doGet (HttpServletRequest request, HttpServletResponse response)</li><li>- void doPost (HttpServletRequest request, HttpServletResponse response)</li></ul>

Tabla 42: Clase CL-01

CL-02	
<b>Nombre de clase</b>	CategoryRequestHandler
<b>Descripción</b>	Esta clase es una de las que se encarga de comunicarse con las capas de Modelo y Vista. Y su función es la de consultar, recibir la información del Modelo y presentarla en la Vista mediante Categoria.jsp
<b>Atributos</b>	<ul style="list-style-type: none"><li>- String [ ] categorias</li></ul>
<b>Métodos</b>	<ul style="list-style-type: none"><li>- String handleRequest (HttpServletRequest request, HttpServletResponse response)</li></ul>

Tabla 43: Clase CL-02

CL-03	
<b>Nombre de clase</b>	AdvancedSearchRequestHandler
<b>Descripción</b>	Esta clase es una de las que se encarga de comunicarse con las capas de Modelo y Vista. Y su función es la de procesar la información del usuario, consultar a la capa de Modelo, recibir la información y presentarla en la Vista mediante Inicio.jsp
<b>Atributos</b>	
<b>Métodos</b>	<ul style="list-style-type: none"><li>- String handleRequest (HttpServletRequest request, HttpServletResponse response)</li></ul>

Tabla 44: Clase CL-03

CL-04	
<b>Nombre de clase</b>	QueryRequestHandler
<b>Descripción</b>	Esta clase es una de las que se encarga de comunicarse con las capas de Modelo y Vista. Y su función es la de procesar la información del usuario, consultar a la capa de Modelo, recibir la información y presentarla en la Vista mediante Inicio.jsp
<b>Atributos</b>	
<b>Métodos</b>	- String handleRequest (HttpServletRequest request, HttpServletResponse response)

*Tabla 45: Clase CL-04*

CL-05	
<b>Nombre de clase</b>	RandomDataRequestHandler
<b>Descripción</b>	Esta clase es una de las que se encarga de comunicarse con las capas de Modelo y Vista. Y su función es la de procesar la información para hacer que los resultados estén ordenados aleatoriamente, consultar a la capa de Modelo, recibir la información y presentarla en la Vista mediante Inicio.jsp
<b>Atributos</b>	
<b>Métodos</b>	- String handleRequest (HttpServletRequest request, HttpServletResponse response)

*Tabla 46: Clase CL-05*

<b>CL-06</b>	
<b>Nombre de clase</b>	AddDocumentRequestHandler
<b>Descripción</b>	Esta clase es la encargada de recibir la información que se necesita para añadir un nuevo video al sistema. Entre sus funciones está la de generar los documentos xml necesarios para indexar el nuevo video a los índices de Solr. Además de esto también es la clase encargada de limpiar los subtítulos para usarlos como pide el sistema.
<b>Atributos</b>	- String [ ] campos
<b>Métodos</b>	<ul style="list-style-type: none"> <li>- String handleRequest (HttpServletRequest request, HttpServletResponse response)</li> <li>- void createDocumentXML (String campos [ ])</li> <li>- void createDocumentXMLMT (String campos [ ])</li> <li>- void createDocumentsXMLMT (String campos [ ])</li> <li>- Map&lt;String, String&gt; subtitlesWithTimeMark (String subtítulos)</li> <li>- String clearSubtitlesText (String subtítulos)</li> <li>- boolean stringIsNumber (String cadena)</li> </ul>

*Tabla 47: Clase CL-06*

#### 4.2.2.2 Clases que forman la capa de Modelo

CL-07	
<b>Nombre de clase</b>	SolrQueries
<b>Descripción</b>	Esta clase es la encargada de recibir las llamadas de los RequestHandlers para así saber realizar el tipo de consultas que pide la llamada. Siendo así la clase encargada de devolver toda la información que salga de la base de datos.
<b>Atributos</b>	<ul style="list-style-type: none"> <li>- String [ ] campos</li> <li>- String [ ] camposMT</li> <li>- double numeroDocumentos</li> </ul>
<b>Métodos</b>	<ul style="list-style-type: none"> <li>- SolrQueries ( )</li> <li>- ArrayList&lt;Map&lt;String, String&gt;&gt; DoSolrQuery (SolrQuery query, boolean coreSelector)</li> <li>- ArrayList&lt;Map&lt;String, String&gt;&gt; DoSimpleQuery (String consulta, int maxResultados, int primerResultado)</li> <li>- ArrayList&lt;Map&lt;String, String&gt;&gt; DoRandomQuery (String consulta, int maxResultados, int primerResultado, int seed)</li> <li>- ArrayList&lt;Map&lt;String, String&gt;&gt; DoAdvancedQuery (String [] consulta, int maxResultados, int primerResultado)</li> <li>- double getNumeroDocumentos ( )</li> <li>- void setNumeroDocumentos (double numeroDocumentos)</li> </ul>

Tabla 48: Clase CL-07

CL-08	
<b>Nombre de clase</b>	SolrConex
<b>Descripción</b>	Esta clase es la encargada de realizar la conexión con la base de datos, permitiendo seleccionar el índice que se desea consultar.
<b>Atributos</b>	<ul style="list-style-type: none"> <li>- String core</li> <li>- String coreMT</li> <li>- SolrClient solr</li> </ul>
<b>Métodos</b>	<ul style="list-style-type: none"> <li>- SolrClient getSolr ( )</li> <li>- void setSolr (SolrClient solr)</li> </ul>

Tabla 49: Clase CL-08

### 4.2.3. Diagramas de Secuencia

En este apartado se mostrarán mediante el uso de diagramas de secuencia las interacciones que se dan en el sistema en cada uno de los casos de uso definidos en el apartado 3.2.2 Especificación de los Casos de Uso

- Diagrama de Secuencia del caso de uso CU-01. Obtención de transcripción utilizando la herramienta Speech to Text.

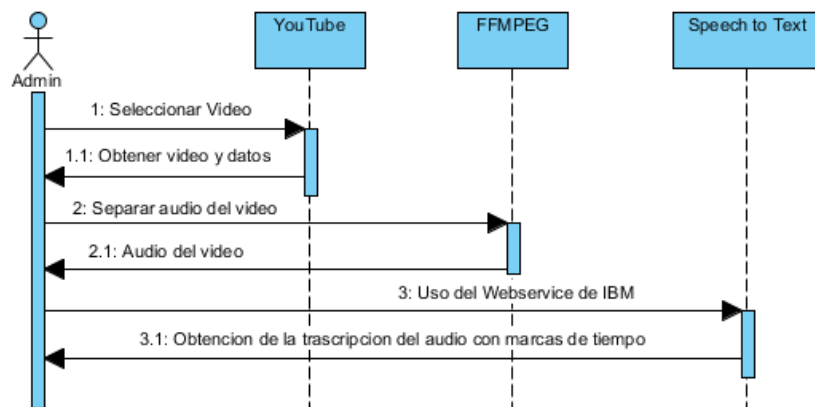


Ilustración 12: Diagrama de secuencia CU-01

- Diagrama de Secuencia del caso de uso CU-02. Añadir nuevo video al índice.

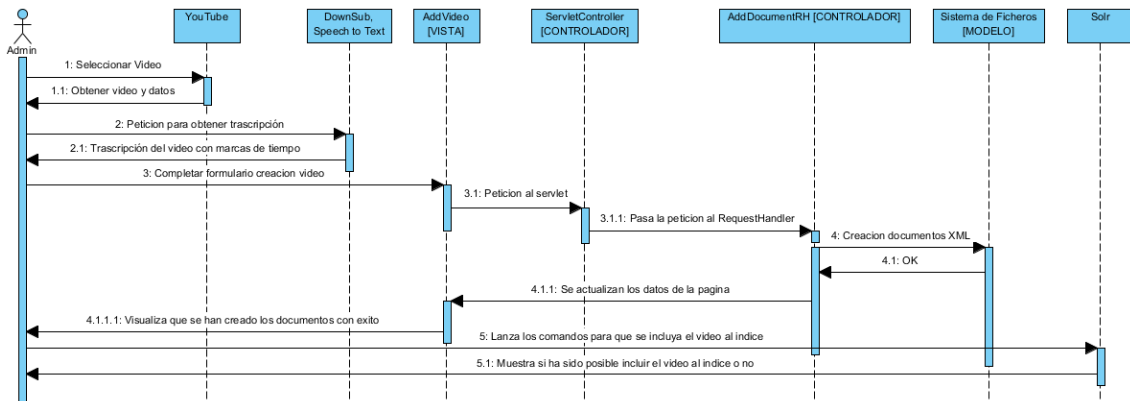


Ilustración 13: Diagrama de secuencia CU-02

- Diagrama de Secuencia de los casos de uso CU-03, CU-04 y CU-05. Estos tres casos de uso tienen prácticamente el mismo diagrama de secuencia, ya que las diferencias que tienen son la del RequestHandler que se encarga de la petición y de la vista a la que se dirigen.

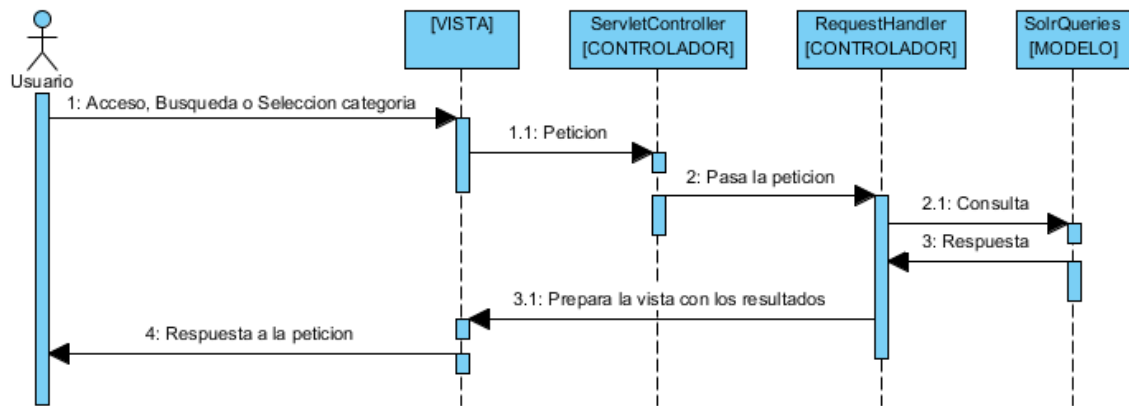


Ilustración 14: Diagrama de secuencia CU-03, CU-04 y CU-05

- Diagrama de Secuencia del caso de uso CU-06. Búsqueda Avanzada.

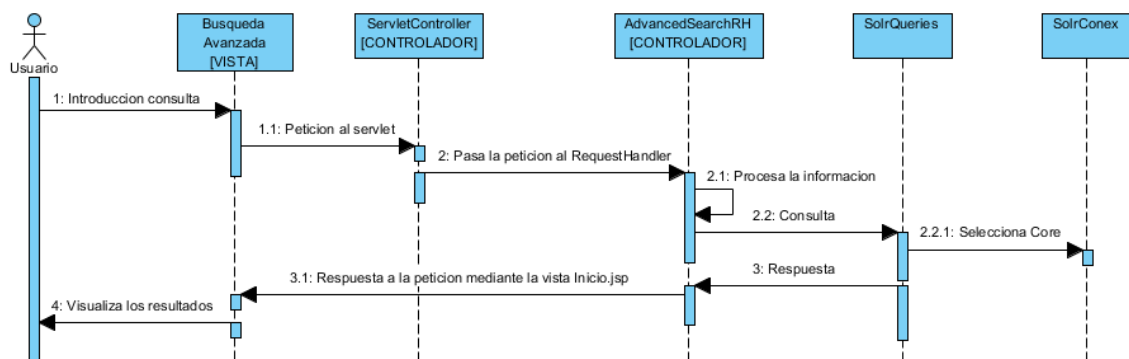


Ilustración 15: Diagrama de secuencia CU-06

## 4.3. Diseño Físico de Datos

El siguiente apartado del diseño del sistema es el del diseño físico de datos, en el que se establecen las estructuras físicas de los datos que utilizara el sistema.

Para ello se tendrán en cuenta los requisitos establecidos en la fase de análisis del sistema, así como las características específicas del entorno tecnológico para tratar los datos como queremos que sean tratados.

### 4.3.1. Diseño del modelo físico de la Base de Datos

Como se ha ido comentando en apartados anteriores, la base de datos que utilizara el sistema será la que incluye el motor de búsquedas, ya que permite almacenar la información además de indexarla o tratarla como se desee. Para ello se han tenido que modificar documentos de configuración de Solr.

Como también se ha comentado, se crearán y utilizaran dos índices sobre el mismo motor de búsquedas, esto se debe a que se pretende ofrecer una búsqueda a nivel de video y otra búsqueda a nivel de localizar en el tiempo del video cuando se menciona lo consultado por el usuario, es por esto por lo que se necesitan dos índices con similitudes, pero con funcionamientos distintos. Los índices en Solr también son conocidos como Cores, por lo que a continuación se pueden utilizar ambas palabras para referirnos a los índices.

#### 4.3.1.1 Índice SubSearch

Este primer índice o core es el encargado de indexar los videos como conjunto, es decir sin tener en cuenta marcas de tiempo. Para ello se ha modificado el documento de configuración del core, este documento se conoce como managed-schema.xml dentro de Solr.

Esta modificación consiste en añadir los siguientes campos al índice, para que así puedan ser consultados y almacenados o no dependiendo de las opciones que se pueden ver en la siguiente ilustración:



```

<field name="id_video" type="string" indexed="true" stored="true" required="true" multiValued="false" />
<field name="user" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="titulo" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="autor" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="fecha_publicacion" type="date" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="categoria" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="descripcion" type="string" indexed="false" required="true" stored="true" multiValued="false"/>
<field name="contenido" type="text_general" indexed="true" required="true" stored="false" multiValued="false"/>

```

*Ilustración 16: Configuración core SubSearch*

Los campos añadidos coinciden con los campos que debe de tener cada uno de los videos, por eso se marca la opción de required como true, ya que no podrá existir ningún video en el sistema que carezca de alguno de los campos. Además de esto destacar la opción de stored que permite almacenar la información para utilizarse como Base de datos, que está en todos los campos a true excepto en el del contenido, ya que son los subtítulos, y con que estén indexados es suficiente para lo que se quiere ofrecer con el sistema.

#### 4.3.1.2 Índice SubSearchMT

Este es el segundo índice del sistema, y es el encargado de indexar a nivel de marcas de tiempo, es decir en las transcripciones de un video se obtienen pares de marca de tiempo y frase dicha en ese tiempo, por esto se necesita almacenar cada una de las frases con la marca de tiempo en la que es dicha, y con esto poder dirigir al video en el momento justo en el que se habla de lo buscado por el usuario.

Para ello se han tomado las mismas modificaciones que con el core anterior, añadiendo un nuevo campo, el de la marca de tiempo, quedando así:

```

<field name="id_video" type="string" indexed="true" stored="true" required="true" multiValued="false" />
<field name="user" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="titulo" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="autor" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="fecha_publicacion" type="date" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="categoria" type="string" indexed="true" required="true" stored="true" multiValued="false"/>
<field name="descripcion" type="string" indexed="false" required="true" stored="true" multiValued="false"/>
<field name="contenido" type="text_general" indexed="true" required="true" stored="false" multiValued="false"/>
<field name="marca_tiempo" type="string" indexed="false" required="true" stored="false" multiValued="false"/>

```

*Ilustración 17: Configuración core SubSearchMT*

### 4.3.2. Diseño del modelo físico del sistema de ficheros

Además de guardar la información de los videos en la Base de Datos, también se utiliza el sistema de ficheros donde se guardarán los documentos necesarios para indexar cada uno de los videos en el índice que le corresponda, para ello se organizan los directorios de la siguiente forma:

- **Directorio raíz:** esta carpeta se ha decidido llamarla Documentos SubSearch, y contendrá los archivos de ambos cores o índices. Dentro se encuentran los siguientes directorios:
  - **Documentos:** En este directorio se almacena cada uno de los documentos generados para indexar el video en el core SubSearch, es decir carecen de marcas de tiempo, por lo que esto quiere decir que un documento es un video en el sistema. Un ejemplo de documento almacenado es el siguiente:

```
<add>
  <doc>
    <field name="id_video">V6zpM3pm8rU</field>
    <field name="user">adcouncil</field>
    <field name="titulo">Children are Children (:60) | Welcoming Refugees | Ad Council</field>
    <field name="autor">Ad Council</field>
    <field name="fecha_publicacion">2016-06-20</field>
    <field name="categoria">#Entertainment</field>
    <field name="descripcion">Refugees flee a life that is threatened daily by war and civil unrest. Visit
      http://www.EmbraceRefugees.org to learn the facts and access resources to help welcome refugees in your
      community. Fore information, news, events, and more, please click here: http://po.st/AdCouncil</field>
    <field name="contenido">
      there's a lot of muddling it rains here
      and it makes it really hard to find food
      your car bombs every day my mom worries
      about me when I go out every time I you
      got alarm bell go off from school i
      think it's an air raid
      sometimes I have nightmares about it a
      lot of houses in her neighborhood have
      been destroyed and like to close my ears
      and sing songs whenever the bombs can
      close my dad says we have to leave it
      makes be scared
      I'm worried our new neighbors were like
      us what if they don't want to sailer
      would because we don't speak the
      language
      it might be hard to make friends but I
      know it's all going to be okay but im
      gonna erase it
      I just want my family to be safe but
      these are not my words are not my order
      these are not my words
      yeah
    </field>
  </doc>
</add>
```

*Ilustración 18: Ejemplo documento core SubSearch*

- **DocumentosMT:** Este directorio contiene un directorio por cada uno de los videos del sistema, esto es porque es el encargado de almacenar los documentos con marcas de tiempo. Es decir, será desde donde se indexen los videos al core SubSearchMT. Cada uno de los directorios que contiene estará identificado por el id del video, y contendrá tantos documentos xml como pares de marca de tiempo/frase contenga en su transcripción. Un ejemplo del documento que almacenan es el siguiente:

```

<add>
  <doc>
    <field name="id_video">aMcK1tS2Zs</field>
    <field name="user">latenight</field>
    <field name="titulo">President Obama Explains His Old-School Blackberry</field>
    <field name="autor">The Tonight Show Starring Jimmy Fallon</field>
    <field name="fecha_publicacion">2016-06-10</field>
    <field name="categoria">#Comedy</field>
    <field name="descripcion">President Obama shares his tips on compromise and breaks down why he doesn't
      carry a modern smartphone. President Obama Explains His Old-School Blackberry http://www.youtube.com/
      fallontonight</field>
    <field name="marca_tiempo">00h00m24s</field>
    <field name="contenido">
      One is despite the news,
      despite all the rancor,
    </field>
  </doc>
</add>

```

Ilustración 19: Ejemplo documento core SubSearchMT

## 4.4. Diseño de interfaces de usuario

En esta sección se definen las interfaces con las que interactuará el usuario una vez la página web se encuentre en funcionamiento.

Las interfaces de usuario de las que dispondrá el sistema son las siguientes:

### 4.4.1. Página principal

Esta página contendrá la lista de videos de los que dispone el sistema ordenados de manera aleatoria, este orden cambiara cada vez que se recargue o se vuelva a acceder a la página principal.

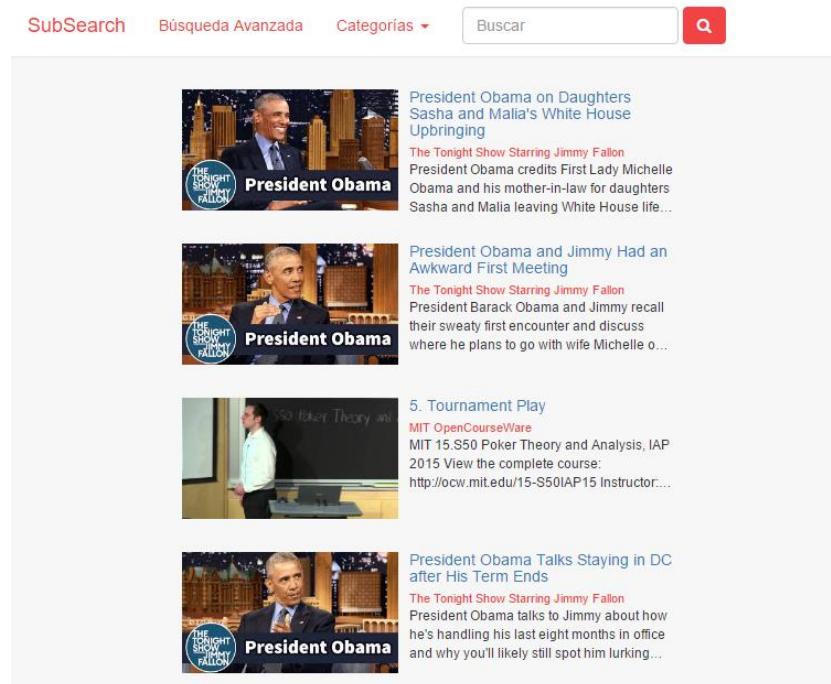
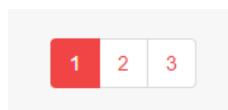


Ilustración 20: Interfaz página principal

Como se puede apreciar, desde la barra de navegación se tiene acceso a todas las funcionalidades que se pretenden ofrecer al usuario. En esta captura además se puede ver cómo quedaría la lista de los resultados, tal y como se dice en los requisitos, apareciendo por pantalla un máximo de 10 resultados, en el caso de que existieran más resultados, la vista realizaría automáticamente la paginación de los mismos. La paginación de resultados destacaría la página en la que se encuentra el usuario en ese mismo instante y permite acceder al resto de resultados al seleccionar la página que se desea. Un ejemplo de cómo se mostraría la paginación es el siguiente:



*Ilustración 21: Interfaz paginación de resultados*

Este estilo de paginación será idéntico para cada una de las paginaciones que realicen las vistas en cualquiera de las paginas en las que se muestran resultados.

El estilo de página principal será el utilizado como página en la que se mostraran los resultados de consultas básicas y avanzadas, en el caso de realizar una consulta, ya sea básica o avanzada y el sistema no encuentre resultados aparecerá el mensaje de que no se han encontrado resultados.

Sobre los resultados obtenidos se seguirá el siguiente estilo:



*Ilustración 22: Interfaz estilo de resultados*

Mostrándose el título del video en color azul y enlazado con el video, al hacer clic el usuario se abrirá en una nueva ventana el video. En el caso de incluir marcas de tiempo en los resultados, el video se abrirá en el momento justo en el que se habla sobre lo consultado.

Por otro lado, aparecerá la miniatura del video que también enlaza con el video de la misma manera que lo hace el título. El ultimo enlace se encuentra en el nombre del canal, marcado en color rojo y enlazando al canal del autor del video.

Finalmente se mostrará un pequeño fragmento de la descripción del video para que el usuario pueda hacerse una idea de lo que trata el video.

#### 4.4.2. Página de Categorías

La página de categorías podrá ser accedida desde el desplegable de la barra de navegación en la sección de categorías.

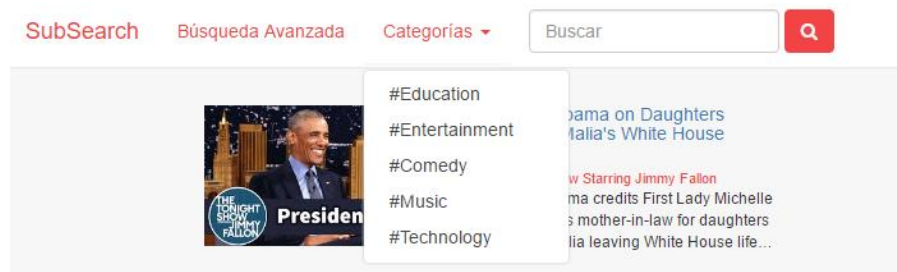


Ilustración 23: Desplegable categorías

Al pulsar en cualquiera de ellas se abrirá la siguiente página, mostrando todos los videos de la categoría seleccionada. Al igual que ocurre con la página principal, mostrara 10 resultados y en el caso de existir más se paginarán.

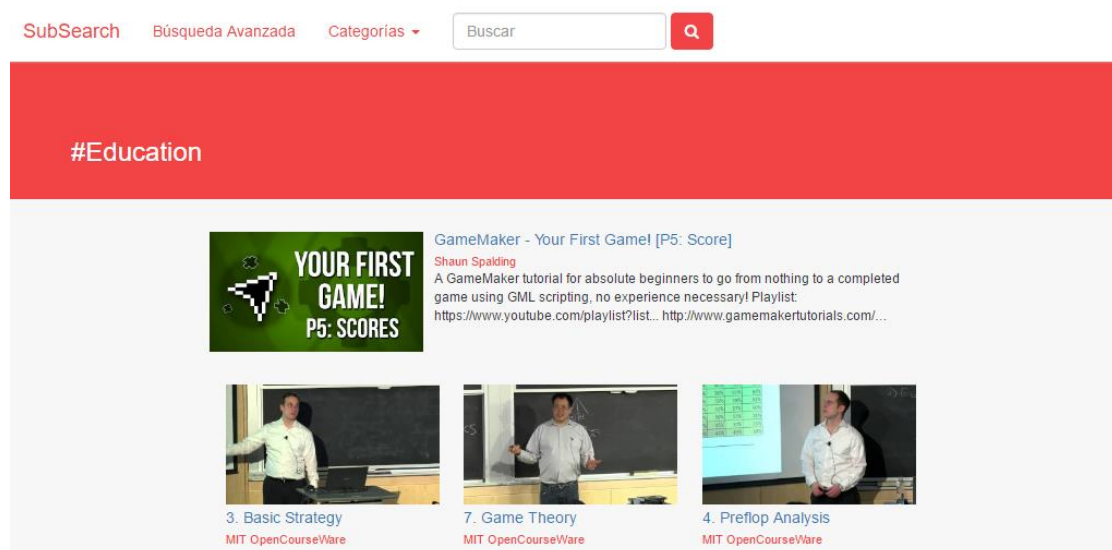
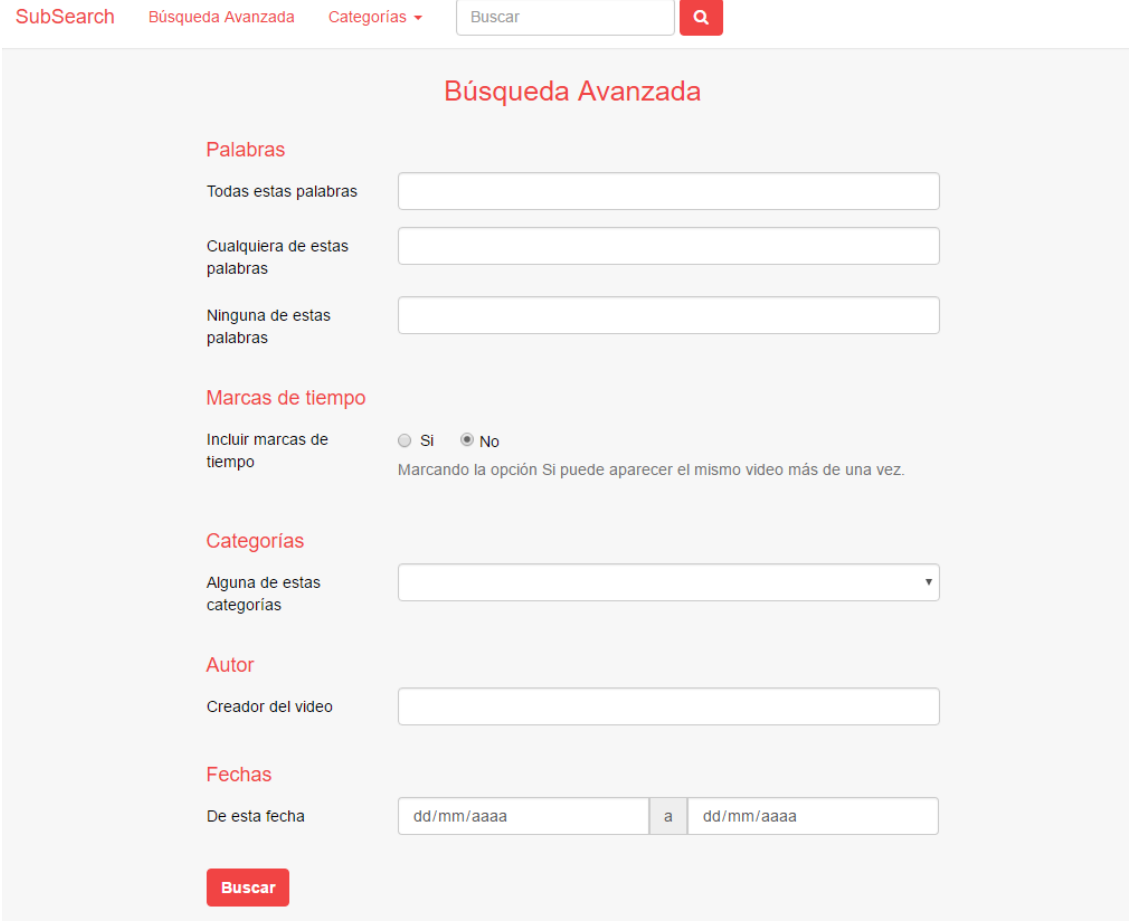


Ilustración 24: Interfaz categorías

De esta página destacar que en la barra roja de la parte de arriba se mostrara la categoría seleccionada para saber en la categoría que se encuentra.

#### 4.4.3. Búsqueda avanzada

La sección de búsqueda avanzada no mostrará ningún resultado por pantalla, ya que se ha decidido que únicamente este formada por un formulario que se encargue de lanzar la consulta, separando la entrada de la consulta con los resultados de la misma. El formulario de consulta es el siguiente:



SubSearch   Búsqueda Avanzada   Categorías ▾  

### Búsqueda Avanzada

**Palabras**

Todas estas palabras

Cualquiera de estas palabras

Ninguna de estas palabras

**Marcas de tiempo**

Incluir marcas de tiempo   ☐ Si   ☒ No

Marcando la opción Si puede aparecer el mismo video más de una vez.

**Categorías**

Alguna de estas categorías

**Autor**

Creador del video

**Fechas**

De esta fecha    a

*Ilustración 25: Formulario búsqueda avanzada*

En él se pueden diferenciar las diferentes secciones que tiene, pudiéndose estas combinarse a la vez tal y como desee el usuario.

Las secciones por las que se forma son:

- **Palabras:** sección en la que se introducirán o no las palabras clave que se pretenden encontrar o no.

- Todas estas palabras: este campo de entrada de texto buscara todas aquellas palabras que sean incluidas, mostrando los resultados que las contengan.
- Cualquiera de estas palabras: este campo permite introducir palabras que pueden ser clave pero que el usuario no está seguro de ello, por lo que se pueden encontrar resultados que las contengan o no.
- Ninguna de estas palabras: este campo permite introducir palabras que no se desean que aparezcan en los resultados.
- **Marcas de tiempo:** Esta sección, como ya se ha comentado en apartados anteriores, es la encargada de seleccionar el índice utilizado por el motor de búsqueda, pudiendo seleccionar entre el motor que incluye las marcas de tiempo y el motor que no las incluye. Avisando en todo momento al usuario que si se marca la opción de buscar incluyendo marcas de tiempo puede aparecer el mismo video varias veces, al tratarse como unidad la marca de tiempo y no el video. La opción por defecto es la de realizar la búsqueda sin marcas de tiempo.
- **Categorías:** Esta sección es la encargada de seleccionar la categoría de la que se desean obtener los resultados. Está formada por una lista desplegable con cada una de las categorías que se encuentran en el sistema, además incluye un campo vacío que es el que viene por defecto para poder encontrar resultados de cualquiera categoría.
- **Autor:** Esta sección permite al usuario buscar algún autor de un video, ya sea por el nombre del canal o el nombre de usuario.
- **Fechas:** Por último, la sección de fechas permite seleccionar los videos que se hayan subido antes, después y entre dos fechas, mediante el uso de dos calendarios despegables.

Además de esto destacar que el botón de búsqueda permanecerá desactivado si no se han introducido los suficientes datos de entrada como para lanzar la consulta.

## 4.5. Especificación del entorno de desarrollo

En este apartado se enumerarán las diferentes herramientas que se han utilizado durante el desarrollo del sistema. Estas herramientas han sido seleccionadas tras tomar decisiones como el lenguaje de programación en el que se va a realizar, software que se va a utilizar para facilitar la comunicación entre los distintos módulos, y el poder replicar el proyecto en diferentes sistemas operativos. Para ello se enumerarán y explicarán las decisiones tomadas en los siguientes aspectos:



#### 4.5.1. Hardware

Los equipos informáticos de los que se ha dispuesto para el desarrollo del proyecto, ya sea para su creación, como para la realización de pruebas, son los siguientes:

<b>Modelo</b>	MacBook Pro Retina 13"
<b>Procesador</b>	2,7GHz Intel Core i5
<b>Memoria RAM</b>	8 GB 1867 MHz DDR3
<b>Disco Duro</b>	250GB SSD
<b>Tipo de Sistema</b>	Sistema operativo de 64 bits, procesador x64
<b>Sistema Operativo</b>	OS X El Capitan Versión 10.11.6

*Tabla 50: Configuración ordenador 1*

<b>Modelo</b>	Personal Computer
<b>Procesador</b>	AMD FX Series FX-8350 4.0Ghz 8X Black Edition
<b>Memoria RAM</b>	8 GB 1866 MHz DDR3
<b>Disco Duro</b>	1TB HDD
<b>Tipo de Sistema</b>	Sistema operativo de 64 bits, procesador x64
<b>Sistema Operativo</b>	Windows 10

*Tabla 51: Configuración ordenador 2*

<b>Modelo</b>	iPhone 6s
<b>Procesador</b>	1,85GHz 2X Apple A9
<b>Memoria RAM</b>	2 GB DDR4
<b>Memoria</b>	16GB
<b>Tipo de Sistema</b>	Sistema operativo móvil de 64 bits, procesador x64
<b>Sistema Operativo</b>	iOS 9.3.5

*Tabla 52: Configuración dispositivo móvil*

#### 4.5.2. Sistemas Operativos

Durante el desarrollo del proyecto se han utilizado varios sistemas operativos distintos que se exponen en este apartado, detallando las funcionalidades por las que se han utilizado cada uno de ellos. La elección de estos sistemas operativos se basa en la disponibilidad, conocimientos de uso, compatibilidad con los elementos del entorno de desarrollo utilizados en el proyecto y pruebas de funcionamiento en dispositivos móviles.

##### 4.5.2.1 OS X El Capitan

OS X es un entorno operativo basado en Unix, desarrollado, comercializado y vendido por Apple que se incluye en toda su gama de ordenadores. Este es el sistema en el que se ha desarrollado la gran parte de la aplicación debido a la facilidad de uso de



herramientas básicas para el sistema mediante la consola. Además de ser la única opción de sistema operativo debido al hardware disponible por el alumno que permitiera la movilidad requerida en algunos momentos del desarrollo.

Este sistema operativo al igual que otros posee una gran variedad de aplicaciones destinadas a los desarrolladores, presentando un entorno gráfico minimalista e intuitivo que facilita su uso. Además, utiliza una consola muy similar a la de sistemas GNU/Linux, por lo que se puede aplicar lo aprendido durante los cursos universitarios.

#### 4.5.2.2 Windows 10

Windows 10 es un sistema operativo desarrollado por Microsoft que junto a otras versiones más antiguas son los sistemas operativos más conocidos y utilizados en el mundo. Este sistema operativo puede instalarse en cualquier dispositivo compatible, y necesita poseer una licencia de pago. Este sistema operativo ha sido utilizado con determinadas aplicaciones que no existían en OS X y no había alternativa a utilizar herramientas similares al no existir apenas. El sistema también se ha utilizado a la hora del desarrollo de la interfaz gráfica al disponer de mejores pantallas para el tema de estilos y tamaños de los elementos de la página.

#### 4.5.2.3 iOS 9.3.5

iOS es el sistema operativo móvil de Apple que incluyen los iPhone, iPad y iPod. La utilización de un sistema operativo móvil se debe principalmente para realizar comprobaciones de cómo se vería la interfaz gráfica del sistema en este tipo de dispositivos y comprobar su correcto funcionamiento en pantallas táctiles de tamaño reducido.

### 4.5.3. Lenguajes de programación

En este apartado se enunciarán los lenguajes de programación que se utilizan en proyectos de aplicaciones J2EE. Además, se incluirá el lenguaje utilizado en el desarrollo de la herramienta Speech to Text.

#### 4.5.3.1 Java

Java es un lenguaje de programación de propósito general, concurrente, orientado a objetos que fue diseñado por Sun Microsystems (adquirida por la compañía Oracle) [15]. Gracias a la filosofía de diseño de Java, conocida como WORA (“Write once, run

anywhere”) permite escribir una vez el código y ejecutarlo en cualquier dispositivo, esto favoreció a su crecimiento como lenguaje de programación al ofrecer una gran amplitud de uso en distintos ámbitos de la industria de la informática, como pueden ser:

- **En dispositivos móviles y sistemas embebidos:** Desde la creación de la especificación J2ME (Java 2 Platform, Micro Edition).
- **En el navegador web:** Posibilidad de desarrollar pequeñas aplicaciones conocidas como Applets en Java que pueden ser incrustadas en páginas HTML.
- **En sistemas de servidor:** Entorno de funcionamiento más popular para Java, desde la aparición de la especificación de Servlets y JSP (Java Server Pages). Por este aspecto es por el que se ha decidido seleccionar a Java como el lenguaje de programación del sistema.
- **En aplicaciones de escritorio:** Con la aparición de bibliotecas de desarrollo gráfico y de que el entorno de ejecución Java conocido como JRE se haya convertido en un componente habitual de los PC de usuario.
- **Plataformas soportadas:** Existe una versión del JRE para la mayoría de equipos de escritorio. En los sistemas de Microsoft no está incluido por defecto. En el caso del sistema operativo Mac OS, incluye una versión propia del JRE, al igual que ocurre con la mayoría de distribuciones de GNU/Linux.

Otro de los motivos por el que se ha seleccionado Java como lenguaje de programación es que ya ha sido utilizado por el alumno en anteriores proyectos, por lo que se dispone de experiencia, además de conocer también Eclipse, el entorno de desarrollo seleccionado.

Además, hay que destacar que se utilizará el motor de búsquedas Solr que está basado en la biblioteca Java del proyecto Lucene, por lo que ofrece soporte para Java con su cliente conocido como SolrJ. Ofreciendo una interfaz Java para añadir, actualizar y consultar índices de Solr. [16]

#### 4.5.3.2 Ruby

Ruby es un lenguaje de programación interpretado, reflexivo y orientado a objetos, creado por el programador japonés Yukihiro "Matz" Matsumoto, quien comenzó a trabajar en Ruby en 1993, y lo presentó públicamente en 1995. Combina una sintaxis inspirada en Python y Perl con características de programación orientada a objetos

similares a Smalltalk. Ruby es un lenguaje de programación interpretado en una sola pasada y su implementación oficial es distribuida bajo una licencia de software libre [17].

La decisión de utilizar este lenguaje de programación se ha tomado al encontrarse un ejemplo de uso del webservice de IBM que permite realizar y recibir llamadas, y con esto poder utilizar su sistema de reconocimiento de texto Speech to Text.

#### 4.5.3.3 JavaScript

JavaScript es un lenguaje de programación interpretado, dialecto del estándar ECMAScript. Se define como orientado a objetos, basado en prototipos, imperativo, débilmente tipado y dinámico [18].

Se utiliza principalmente en su forma del lado del cliente, implementado como parte de un navegador web permitiendo mejoras en la interfaz de usuario y páginas web dinámicas, aunque existe una forma de JavaScript del lado del servidor [15].

Tradicionalmente se venía utilizando en páginas web HTML para realizar operaciones y únicamente en el marco de la aplicación cliente, sin acceso a funciones del servidor.

Es por este motivo por el que se ha tenido que utilizar en el sistema, a la hora de que los usuarios del sistema completen los formularios de los que dispone, como son la búsqueda avanzada y el formulario de añadir un nuevo video.

#### 4.5.4. Lenguajes de marcas

Al tratarse del desarrollo de una página web se manejarán lenguajes que no son considerados de programación, es el caso de los lenguajes de marcas, enunciados en este apartado al ser utilizados durante el desarrollo del sistema.

##### 4.5.4.1 HTML

Las siglas HTML del inglés *HyperText Markup Language* (lenguaje de marcas de hipertexto) hace referencia al lenguaje de marcado para la elaboración de páginas web. Es un estándar que sirve de referencia del software que conecta con la elaboración de páginas web en sus diferentes versiones, define una estructura básica y un código (denominado código HTML) para la definición de contenido de una página web, como texto, imágenes, videos, entre otros. Es un estándar a cargo del *World Wide Web*

*Consortium* (W3C), organización dedicada a la estandarización de las tecnologías ligadas a la web. Es el estándar que se ha impuesto en la visualización de páginas web y es el que todos los navegadores actuales han adoptado.

En este proyecto se ha combinado el uso de HTML con Java, esto es posible gracias a la tecnología JSP (JavaServer Pages) que ayuda a crear páginas web que mezclan componentes estáticos con dinámicos, siendo básicamente una página Web con HTML tradicional y código Java incrustado.

La extensión de fichero de una página JSP es ".jsp" en vez de ".html" o ".htm", y eso le indica al servidor que esta página requiere un tratamiento especial, que se puede resumir en que cuando un cliente solicita una página jsp, se ejecuta en el servidor el código JSP de la página, dando como resultado una página HTML que se fusiona con el HTML original, generando una página HTML de respuesta que será enviada al cliente.

#### 4.5.4.2 CSS

CSS son las siglas en inglés de *Cascading Style Sheets*, o en castellano de hojas de estilo en cascada. Es un lenguaje usado para definir y crear la presentación de un documento estructurado escrito en HTML. El *World Wide Web Consortium* (W3C) es el encargado de formular la especificación de las hojas de estilo que servirán de estándar para los agentes de usuario o navegadores.

La idea que se encuentra detrás del desarrollo de CSS es separar la estructura de un documento de su presentación. La información de estilo puede ser definida en un documento separado o en el mismo documento HTML.

Este lenguaje es necesario a la hora de crear el estilo de la interfaz de usuario, como pueden ser los tamaños de los elementos, tipo de letra, tamaño de letra, paleta de colores de la aplicación web...

#### 4.5.4.3 XML

XML viene del inglés *eXtensible Markup Language*, y se trata de un meta-lenguaje que permite definir lenguajes de marcas desarrollado por el *World Wide Web Consortium* (W3C) utilizado para almacenar datos en forma legible. Proviene del lenguaje SGML y permite definir la gramática de lenguajes específicos para estructurar documentos

grandes. A diferencia de otros lenguajes, XML da soporte a bases de datos, siendo útil cuando varias aplicaciones deben comunicarse entre sí o integrar información.

Es una tecnología sencilla que tiene a su alrededor otras que la complementan y la hacen mucho más grande y con unas posibilidades mucho mayores. Tiene un papel muy importante en la actualidad ya que permite la compatibilidad entre sistemas para compartir la información de una manera segura, fiable y fácil.

Por este motivo se utiliza el lenguaje en este proyecto, ya que la definición de los documentos (vídeos) que serán la unidad de trabajo de la aplicación web está hecha en formato ".xml". Además de que existen varios documentos de configuración de herramientas utilizadas que necesitan conocimientos de xml para definir nuevos parámetros, conseguir las dependencias de módulos y componentes externos utilizando MAVEN...

#### 4.5.5. Software de desarrollo

En este apartado se incluye todo tipo de software utilizado durante el desarrollo del proyecto, esto quiere decir que se incluyen tanto aplicaciones de escritorio como páginas web que ofrecen servicios que han servido de utilidad.

##### 4.5.5.1 Eclipse

Eclipse es una plataforma de software compuesto por un conjunto de herramientas de programación de código abierto multiplataforma para desarrollar aplicaciones. Esta plataforma, típicamente ha sido usada para desarrollar entornos IDE (*Integrated Development Environment*), como el IDE de Java llamado Java Development Toolkit (JDT) y el compilador (ECJ) que se entrega como parte de Eclipse.

En este caso se utilizará el Eclipse IDE for Java EE Developers, necesario para crear herramientas cliente-servidor. La versión de Eclipse a utilizar es la conocida como Eclipse Mars 2 [19]. Teniendo en cuenta el añadido del servidor en el que se cargara el sistema, ya que Eclipse permite instalar varios servidores, se ha decidido utilizar el servidor Apache Tomcat v8.0.

#### 4.5.5.2 Solr

Solr es un motor de búsqueda de código abierto basado en la biblioteca Java del proyecto Lucene, con APIs en XML/HTTP y JSON, resaltado de resultados, búsqueda por facetas, caché, y una interfaz para su administración [11]. Corre sobre un contenedor de servlets Java como Apache Tomcat.

Este será el motor de búsqueda utilizado como ya se ha comentado en apartados anteriores, siendo la razón de esta decisión el que se trata de un motor de código abierto, por lo que puede ser utilizado de forma gratuita. Además, en su página web ofrece varios tutoriales para iniciarse, una amplia documentación en la que se puede consultar cualquier duda, al igual que la comunidad que tiene detrás.

#### 4.5.5.3 Sublime Text, Notepad++, Xcode

Estos tres programas han sido utilizados como editores de texto a la hora de modificar, crear y visualizar documentos ".xml", ".html" y ".css". Ambos son gratuitos, aunque Sublime Text es gratuito en su versión de prueba, ya que también dispone de una versión con licencia, por lo que no es un programa de código abierto, al igual que ocurre con Xcode [20]. En cambio, Notepad++ es la única de las tres que es de código abierto.

De las tres aplicaciones, solo Notepad++ [21] y Sublime Text [22] pueden ser usadas tanto en Windows como en OS X, esto no ocurre con Xcode que solo puede usarse en equipos con OS X, al tratarse del entorno de programación que ofrece Apple para el desarrollo de sus aplicaciones.

Además, destacar que Sublime ha sido el editor de texto utilizado a la hora de programar la herramienta de Speech to Text, ya que ofrece plugins para el lenguaje Ruby, haciendo más sencilla la programación en este lenguaje.

#### 4.5.5.4 Google2SRT, DownSub

Estas dos herramientas son vitales para el proyecto, ya que los subtítulos de los videos ha sido posible obtenerlos gracias a ellas.

Google2SRT [23] se trata de una aplicación de escritorio que únicamente se encuentra disponible para el sistema operativo Windows, y permite convertir subtítulos de videos de YouTube al formato SubRip (".srt") conocido por la mayoría de reproductores de

video. El funcionamiento de la herramienta es muy simple, solo basta con introducir el enlace del video del que queremos obtener los subtítulos, la aplicación buscará los subtítulos disponibles para el video y seleccionando el idioma deseado se descargará el archivo ".srt" obteniendo los subtítulos con marcas de tiempo.

DownSub [24] se trata de una herramienta que funciona igual que la anterior, pero está en cambio ofrece sus servicios mediante su web, siendo totalmente compatible con cualquier sistema operativo al utilizarse desde el navegador web. Ha sido la aplicación que se ha utilizado la gran mayoría de veces, debido a la ventaja de no tener que cambiar de equipo, pero ha habido momentos durante el desarrollo que no se encontraba disponible, teniendo que utilizar de forma obligada Google2SRT.

#### 4.5.5.5 Google Chrome y navegadores web

Google Chrome [25] es el navegador web que se ha utilizado durante todo el desarrollo del proyecto, tanto para las herramientas web como DownSub como para probar el funcionamiento de la aplicación, diseñar el estilo de la página, etc. Además de este navegador se han utilizado otros como Safari [26] (tanto para iOS como para OS X) y Mozilla Firefox [27], pero únicamente han sido utilizados para comprobar la visualización de la web.

Para ello además del navegador se ha utilizado la extensión Firebug Lite [28] que permite editar elementos de la página web para así probar distintos diseños, depurar errores, ver parámetros enviados al servidor, etc.

#### 4.5.5.6 W3C Markup Validation Service

Esta herramienta también conocida como W3C Validator [29] permite comprobar la validez del etiquetado de documentos HTML, XML, CSS, ... para así encontrar errores de etiquetado, etiquetas anticuadas, warnings al encontrarse elementos dentro de otros que no son válidos, etc.

Esta herramienta web permite subir documentos ".html", ".xml", ".css" entre otros y mostrar por pantalla las líneas en las que se encuentran los errores para así poder corregirlos de manera más rápida y sencilla al ofrecer la propia herramienta sugerencias para corregir el error. Por esto se ha pasado el validador a todas las plantillas de la página web y hojas de estilo.

#### 4.5.5.7 GitHub Desktop

GitHub Desktop [30] es la aplicación que ofrece la plataforma de desarrollo colaborativo GitHub, para mantener un control de versiones del software desarrollado mediante una interfaz intuitiva y simple, cuyo funcionamiento es idéntico a utilizar Git por comandos desde la consola, pero con la ventaja de tener el entorno visual que hace que se entienda todo de forma más sencilla.

Se ha decidido utilizar GitHub y no otra plataforma al tratarse de un servicio gratuito, y de poder compartirlo con cualquiera que conozca la dirección del proyecto.



## 5. Plan de Pruebas

En esta sección se abordará el mecanismo que se va a seguir para comprobar que se está cumpliendo con las funcionalidades especificadas en los apartados 3 y 4, Análisis y Diseño del sistema.

De esta forma, para el desarrollo de las diferentes pruebas se tendrán en cuenta los requisitos, casos de uso, arquitectura... ya especificados.

Antes de especificar las pruebas del sistema, se debe definir el alcance y entorno de las pruebas, para conocer que se va a probar y sobre que entorno se van a realizar las pruebas.

### 5.1. Definición del alcance de las pruebas

En esta sección se va a especificar el alcance de las pruebas que se van a realizar, es decir los diferentes tipos de pruebas a las que se va a someter al sistema. Los tipos de pruebas que se van a realizar son los siguientes:

- **Pruebas Unitarias:** Se especificará la batería de pruebas unitarias del sistema. Especificando pruebas únicamente de caja negra. Estas pruebas se encargarán de probar algunos uno de los métodos de mayor importancia en el sistema por separado.
- **Pruebas de sistema:** Estas pruebas consisten en testear todo el sistema. Las pruebas se realizarán sobre todo el sistema completo, bajo el entorno de pruebas definido.

Estas diferentes pruebas no siempre se realizarán sobre todos los elementos del sistema, quedando siempre constancia del elemento sobre el que se hace la prueba.

### 5.2. Entorno de pruebas

En esta sección se van a enunciar el entorno donde se van a realizar las pruebas. Teniendo en cuenta los recursos software y hardware de los equipos utilizados.

- **Equipo 1:** Apple MacBook Pro
  - Procesador: 2,7GHz Intel Core i5
  - Memoria RAM: 8 GB 1867 MHz DDR3
  - Sistema Operativo: OS X El Capitan Versión 10.11.6
  - Base de Datos/Motor de búsquedas instalada: Solr 6.0.0

- Servidor Web: Apache Tomcat v8.0
- **Equipo 2:** Ordenador de Sobremesa
  - Procesador: AMD FX Series FX-8350 4.0Ghz 8X Black Edition
  - Memoria RAM: 8 GB 1866 MHz DDR3
  - Sistema Operativo: Windows 10
- **Equipo 3:** iPhone 6s
  - Procesador: 1,85GHz 2X Apple A9
  - Memoria RAM: 2 GB DDR4
  - Sistema Operativo: iOS 9.3.5

### 5.3. Especificación de pruebas

#### 5.3.1. Pruebas unitarias de caja negra

En este apartado se describirán las pruebas unitarias de caja negra que serán las encargadas de comprobar que realizando las inserciones sobre los parámetros de entrada de los métodos obtenemos los resultados esperados cumpliendo así el método con su correcta funcionalidad. Estas pruebas se representarán con la siguiente tabla:

PUCN-NN	
<b>Título</b>	
<b>Objetivo</b>	
<b>Clase</b>	
<b>Método</b>	
<b>Implementación</b>	
<b>Resultado</b>	

*Tabla 53: Ejemplo tabla definición prueba unitaria*

Siendo sus campos los siguientes:

- **PUCN-NN:** código identificativo, unívoco, de cada una de las pruebas unitarias de caja negra. La nomenclatura utilizada será: **PUCN-NN**, donde NN es el número que representa a la prueba.
- **Título:** Título que hace referencia a una breve descripción de la prueba.
- **Objetivo:** Objetivo que se busca conseguir con la prueba.

- **Clase:** nombre de la clase que contiene el método de la prueba.
- **Método:** método sobre el que se realiza la prueba.
- **Implementación:** Indica los parámetros introducidos para la realización de la prueba.
- **Resultado:** Es el resultado esperado para los parámetros introducidos en el campo implementación.

A continuación, se identifican las pruebas de caja negra definidas para el sistema:

PUCN-01	
<b>Título</b>	Limpiar marcas de tiempo de subtítulos
<b>Objetivo</b>	Comprobar que el método quita las marcas de tiempo de los subtítulos, devolviendo únicamente el texto.
<b>Clase</b>	AddDocumentRequestHandler
<b>Método</b>	ClearSubtitlesText
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Introducir un String que contenga únicamente marcas de tiempo.</li> <li>2. Introducir un String que contenga marcas de tiempo con frases</li> <li>3. Introducir un String que contenga únicamente texto.</li> </ol>
<b>Resultado</b>	<ol style="list-style-type: none"> <li>1. Devolver un String vacío.</li> <li>2. Devolver un String con las frases.</li> <li>3. Devolver el mismo String</li> </ol>

*Tabla 54: Prueba unitaria PUCN-01*

<b>PUCN-02</b>	
<b>Título</b>	Obtener par <Marca de tiempo, Frase>
<b>Objetivo</b>	Comprobar que el método devuelve en un Map<String, String> el par marca de tiempo/Frase.
<b>Clase</b>	AddDocumentRequestHandler
<b>Método</b>	SubtitlesWithTimeMark
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Introducir un String que contenga únicamente marcas de tiempo.</li> <li>2. Introducir un String que contenga marcas de tiempo con frases.</li> <li>3. Introducir un String que contenga únicamente texto.</li> </ol>
<b>Resultado</b>	<ol style="list-style-type: none"> <li>1. Devolver un Map con las claves siendo cada una de las marcas de tiempo y los valores vacíos.</li> <li>2. Devolver un Map con las claves siendo cada una de las marcas de tiempo y los valores cada una de las frases que le corresponden.</li> <li>3. Devolver null.</li> </ol>

*Tabla 55: Prueba unitaria PUCN-02*

<b>PUCN-03</b>	
<b>Título</b>	Creación documentos XML
<b>Objetivo</b>	Comprobar que el método crea el documento correspondiente al video que se quiere añadir.
<b>Clase</b>	AddDocumentRequestHandler
<b>Método</b>	CreateDocumentXML
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Introducir los parámetros necesarios.</li> </ol>
<b>Resultado</b>	<ol style="list-style-type: none"> <li>1. Documento creado en la ruta correspondiente.</li> </ol>

*Tabla 56: Prueba unitaria PUCN-03*

<b>PUCN-04</b>	
<b>Título</b>	Creación documentos XMLMT
<b>Objetivo</b>	Comprobar que el método crea los documentos correspondientes al video con marcas de tiempo que se quiere añadir.
<b>Clase</b>	AddDocumentRequestHandler
<b>Método</b>	CreateDocumentXMLMT
<b>Implementación</b>	1. Introducir los parámetros necesarios.
<b>Resultado</b>	1. Documentos creados en la ruta correspondiente.

*Tabla 57: Prueba unitaria PUCN-04*

<b>PUCN-05</b>	
<b>Título</b>	Consulta de query a Base de Datos
<b>Objetivo</b>	Comprobar que el método consulta en la base de datos la query que ha recibido por parámetros y devuelve los resultados.
<b>Clase</b>	SolrQueries
<b>Método</b>	DoSolrQuery
<b>Implementación</b>	1. Introducir query y seleccionar core SubSearch. 2. Introducir query y seleccionar core SubSearchMT.
<b>Resultado</b>	1. Devolver resultados sin marcas de tiempo. 2. Devolver resultados con marcas de tiempo.

*Tabla 58: Prueba unitaria PUCN-05*

### 5.3.2. Pruebas de sistema

Las pruebas del sistema tienen como objetivo verificar el correcto funcionamiento del sistema en su conjunto, por lo que se va a probar las funcionalidades que ofrece la interfaz de usuario, como si de un usuario normal se tratase. Para definir estas pruebas se utilizará la siguiente tabla:

PS-NN	
Nombre	
Objetivo	
Implementación	
Resultado	

*Tabla 59: Ejemplo tabla definición prueba de sistema*

Siendo sus campos los siguientes:

- **PS-NN:** código identificativo, unívoco, de cada una de las pruebas de sistema. La nomenclatura utilizada será: **PS-NN**, donde NN es el número que representa a la prueba.
- **Nombre:** nombre que hace referencia a una breve descripción de la prueba.
- **Objetivo:** Objetivo que se busca conseguir con la prueba.
- **Implementación:** describe el desarrollo seguido para la realización de la prueba.
- **Resultado:** Es el resultado esperado tras realizar los pasos de la implementación.

A continuación, se identifican las pruebas de sistema:

<b>PS-01</b>	
<b>Nombre</b>	Acceso a la página web
<b>Objetivo</b>	Demostrar que al acceder a la web se accede a la página de inicio y se muestran resultados de los videos con los que cuenta el sistema ordenados aleatoriamente.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Observar los resultados obtenidos.</li> <li>3. Recargar la página web.</li> <li>4. Observar el cambio de orden de los resultados</li> </ol>
<b>Resultado</b>	Se ordenarán los resultados de la página de inicio al acceder directamente a ella mediante el uso de una seed aleatoria generada al recargar la página.

*Tabla 60: Prueba de sistema PS-01*

<b>PS-02</b>	
<b>Nombre</b>	Búsqueda desde la barra de navegación
<b>Objetivo</b>	Demostrar que al utilizar el campo de búsqueda de la barra de navegación se obtienen resultados relevantes con las palabras introducidas.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Introducir palabras clave a buscar.</li> <li>3. El sistema mostrara los resultados obtenidos en la pantalla de inicio.</li> <li>4. Observar la relevancia de los mismos.</li> </ol>
<b>Resultado</b>	Se buscará en cada uno de los campos que forman los videos las palabras clave introducidas, mostrando los resultados ordenados de mayor a menor relevancia según lo suponga el motor de búsqueda.

*Tabla 61: Prueba de sistema PS-02*

<b>PS-03</b>	
<b>Nombre</b>	Uso de la paginación de resultados
<b>Objetivo</b>	Demostrar que la paginación de resultados funciona como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Introducir palabras clave a buscar.</li> <li>3. El sistema mostrara los resultados obtenidos en la pantalla de inicio, y si la consulta lo requiere realizara la paginación.</li> <li>4. Seleccionar la página a la que se quiere ir.</li> <li>5. Realizar varios cambios de página.</li> <li>6. Observar que los resultados se mantienen siempre en la página en la que se encuentran.</li> </ol>
<b>Resultado</b>	Los resultados se mantienen en la página en la que se encuentran, y no cambian los resultados al cambiar de página.

*Tabla 62: Prueba de sistema PS-03*

<b>PS-04</b>	
<b>Nombre</b>	Acceso a la sección de Categorías
<b>Objetivo</b>	Demostrar que el acceso a la sección de categorías separa realmente los videos en las categorías que existen, y no se cambian de categoría.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Seleccionar una categoría.</li> <li>3. El sistema mostrara los resultados obtenidos en la pantalla de categorías.</li> <li>4. Seleccionar otra categoría distinta.</li> <li>5. Observar que los resultados se mantienen siempre en la misma categoría a la que pertenecen, además de aparecer ordenados de forma aleatoria.</li> </ol>
<b>Resultado</b>	Se obtendrán los videos de la categoría seleccionada, cambiando únicamente el orden de los mismos, ya que se ordenan de forma aleatoria.

*Tabla 63: Prueba de sistema PS-04*



<b>PS-05</b>	
<b>Nombre</b>	Acceso a la página de Búsqueda Avanzada
<b>Objetivo</b>	Demostrar que el acceso a la búsqueda avanzada se realiza como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Observar que se muestra el formulario de consulta.</li> </ol>
<b>Resultado</b>	El sistema mostrara la página de BusquedaAvanzada

*Tabla 64: Prueba de sistema PS-05*

<b>PS-06</b>	
<b>Nombre</b>	Búsqueda de todas estas palabras
<b>Objetivo</b>	Demostrar que el formulario de búsqueda avanzada trata la información que le proporciona el usuario como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Introducir palabras clave que se desean de buscar en el campo de todas estas palabras.</li> <li>4. Lanzar la consulta.</li> <li>5. Observar los resultados obtenidos.</li> </ol>
<b>Resultado</b>	El sistema mostrara en la página de inicio los resultados obtenidos, mostrando los videos que contengan las palabras clave introducidas por el usuario.

*Tabla 65: Prueba de sistema PS-06*

<b>PS-07</b>	
<b>Nombre</b>	Búsqueda de cualquiera de estas palabras
<b>Objetivo</b>	Demostrar que el formulario de búsqueda avanzada trata la información que le proporciona el usuario como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Introducir palabras clave que se desean que aparezcan o no entre los resultados en el campo de cualquiera de estas palabras.</li> <li>4. Lanzar la consulta.</li> <li>5. Observar los resultados obtenidos.</li> </ol>
<b>Resultado</b>	El sistema mostrara en la página de inicio los resultados obtenidos, mostrando videos que contengan o no alguna de las palabras clave introducidas por el usuario.

*Tabla 66: Prueba de sistema PS-07*

<b>PS-08</b>	
<b>Nombre</b>	Búsqueda de ninguna de estas palabras
<b>Objetivo</b>	Demostrar que el formulario de búsqueda avanzada trata la información que le proporciona el usuario como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Introducir palabras clave que se desean que no aparezcan en los resultados en el campo de ninguna de estas palabras.</li> <li>4. Lanzar la consulta.</li> <li>5. Observar los resultados obtenidos.</li> </ol>
<b>Resultado</b>	El sistema mostrara en la página de inicio los resultados obtenidos, mostrando los videos que no contengan las palabras clave introducidas por el usuario.

*Tabla 67: Prueba de sistema PS-08*

<b>PS-09</b>	
<b>Nombre</b>	Búsqueda incluyendo marcas de tiempo
<b>Objetivo</b>	Demostrar que el formulario de búsqueda avanzada trata la información que le proporciona el usuario como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Seleccionar la opción Si en el campo de incluir marcas de tiempo.</li> <li>4. Lanzar la consulta.</li> <li>5. Observar los resultados obtenidos.</li> </ol>
<b>Resultado</b>	El sistema mostrara en la página de inicio los resultados obtenidos, mostrando un mayor número de videos al ser su unidad de trabajo la marca de tiempo.

*Tabla 68: Prueba de sistema PS-09*

<b>PS-10</b>	
<b>Nombre</b>	Búsqueda por categoría
<b>Objetivo</b>	Demostrar que el formulario de búsqueda avanzada trata la información que le proporciona el usuario como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Seleccionar la categoría en la que se quiere buscar en el campo de categoría.</li> <li>4. Lanzar la consulta.</li> <li>5. Observar los resultados obtenidos.</li> </ol>
<b>Resultado</b>	El sistema mostrara en la página de inicio los resultados obtenidos, mostrando los videos que pertenezcan a la categoría seleccionada.

*Tabla 69: Prueba de sistema PS-10*

<b>PS-11</b>	
<b>Nombre</b>	Búsqueda por fechas
<b>Objetivo</b>	Demostrar que el formulario de búsqueda avanzada trata la información que le proporciona el usuario como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Las fechas en las que se desea de buscar en el campo de fechas.</li> <li>4. Lanzar la consulta.</li> <li>5. Observar los resultados obtenidos.</li> </ol>
<b>Resultado</b>	El sistema mostrara en la página de inicio los resultados obtenidos, mostrando los videos que se hayan publicado antes, después o entre las fechas seleccionadas.

*Tabla 70: Prueba de sistema PS-11*

<b>PS-12</b>	
<b>Nombre</b>	Búsqueda avanzada combinando campos
<b>Objetivo</b>	Demostrar que el formulario de búsqueda avanzada trata la información que le proporciona el usuario como debe.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web.</li> <li>2. Acceder a la sección de búsqueda avanzada</li> <li>3. Completar varios campos del formulario.</li> <li>4. Lanzar la consulta.</li> <li>5. Observar los resultados obtenidos.</li> </ol>
<b>Resultado</b>	El sistema mostrara en la página de inicio los resultados obtenidos, mostrando los videos que cumplan las características introducidas por el usuario.

*Tabla 71: Prueba de sistema PS-12*

<b>PS-13</b>	
<b>Nombre</b>	Navegación web desde dispositivo móvil
<b>Objetivo</b>	Demostrar que el sistema utiliza el Framework como se debe y la interfaz de usuario se ajusta al tamaño de pantalla del dispositivo donde se está visualizando.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Acceder a la dirección url de la página web desde un dispositivo móvil.</li> <li>2. Realizar una navegación normal, realizando búsquedas, etc.</li> </ol>
<b>Resultado</b>	La interfaz del sistema se adaptará en todo momento a la pantalla del dispositivo desde el que se está utilizando.

*Tabla 72: Prueba de sistema PS-13*

<b>PS-14</b>	
<b>Nombre</b>	Agregar video al índice
<b>Objetivo</b>	Demostrar que el sistema permite agregar videos al índice y así aumentar la colección.
<b>Implementación</b>	<ol style="list-style-type: none"> <li>1. Ser usuario administrador del sistema.</li> <li>2. Seleccionar un video a añadir.</li> <li>3. Obtener los subtítulos o transcripción.</li> <li>4. Obtener la información del video.</li> <li>5. Completar el formulario.</li> <li>6. Agregar video al índice ejecutando los comandos necesarios.</li> </ol>
<b>Resultado</b>	El sistema generara automáticamente los documentos necesarios para indexar el video que el usuario administrador ha seleccionado.

*Tabla 73: Prueba de sistema PS-14*

## 5.4. Análisis de consistencia

En este apartado se verá reflejado como las pruebas cubren correctamente los requisitos definidos en el apartado 3 de Análisis del sistema.

Para ello se hará uso de una matriz de trazabilidad, Requisitos/Pruebas en la que se podrá comprobar visualmente que la funcionalidad queda verificada con las pruebas realizadas.

	PUCN-01	PUCN-02	PUCN-03	PUCN-04	PUCN-05	PS-01	PS-02	PS-03	PS-04	PS-05	PS-06	PS-07	PS-08	PS-09	PS-10	PS-11	PS-12	PS-13	PS-14
RF-01						X	X	X	X	X	X	X	X	X	X	X	X	X	X
RF-02						X	X	X	X	X	X	X	X	X	X	X	X	X	X
RF-03						X	X		X	X	X	X	X	X	X	X	X	X	X
RF-04						X	X		X	X	X	X	X	X	X	X	X	X	
RF-05						X	X		X	X	X	X	X	X	X	X	X	X	
RF-06					X	X	X		X									X	
RF-07					X					X	X	X	X	X	X	X	X	X	
RF-08					X						X	X	X				X	X	
RF-09				X	X									X			X	X	
RF-10					X										X		X	X	
RF-11					X											X	X	X	
RF-12					X												X	X	
RF-13						X			X									X	
RF-14						X			X									X	
RF-15						X	X		X	X	X	X	X	X	X	X	X	X	
RF-16																			X
RF-17																			X
RF-18			X	X															X
RF-19	X	X																	X
RNF-01						X	X	X	X	X	X	X	X	X	X	X	X	X	
RNF-02						X	X	X	X	X	X	X	X	X	X	X	X	X	
RNF-03						X			X									X	
RNF-04						X			X									X	
RNF-05						X		X	X									X	
RNF-06										X								X	
RNF-07																			X
RNF-08																			X
RNF-09																			X

Tabla 74: Consistencia Requisitos/Pruebas

## 5.5. Estudio de la precisión del sistema

Tras probar el funcionamiento del sistema se pasa a comprobar el funcionamiento real del sistema. Para ello se detallará la técnica utilizada en la valoración de los resultados, además del entorno del sistema.

### 5.5.1. Precisión y Recall

Para la evaluación del sistema se va a utilizar la técnica de Precision-Recall, ya que esta técnica se puede utilizar si se conoce la colección con la que se está evaluando el sistema. En este caso el sistema dispone de una colección de 50 videos que han sido elegidos aleatoriamente tras comprobar la calidad de las transcripciones (La colección de videos viene listada en el Anexo B). Por lo que se conoce al 100% la colección con la que se trabaja, es decir, al hacer cualquier consulta se conocen los resultados que se deberían de obtener si tuviera un funcionamiento perfecto.

La técnica de Precision-Recall busca medir la relevancia de los resultados obtenidos al realizar una consulta determinada, obteniendo la precisión a la hora de listar los resultados. El conjunto de consultas es el siguiente:

- Barack Obama Blackberry
- Graphic card expensive
- How to play with jacks

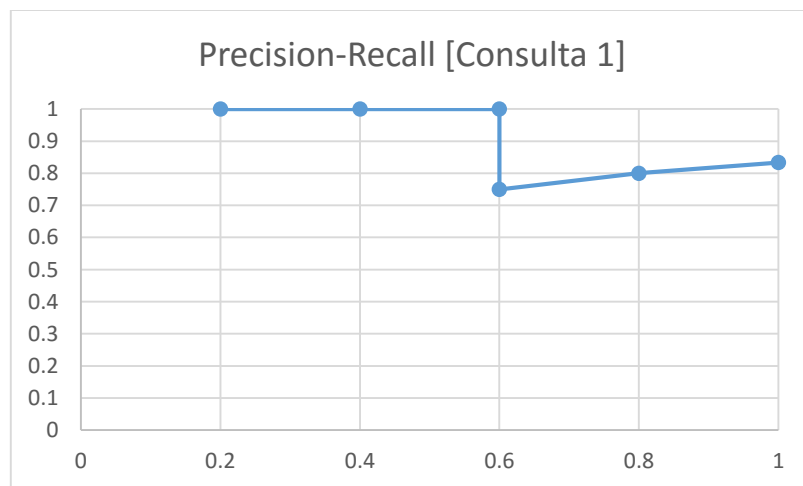
Estas consultas se realizarán mediante el uso de la búsqueda simple, ya que lanzarlas desde la búsqueda avanzada requeriría de más restricciones a la hora de crear la consulta.

La primera consulta realizada es: Barack Obama Blackberry. Ya que se conoce que existen varios videos de Obama en los que habla de su antigua Blackberry. Los resultados obtenidos son los siguientes:

Relevancia	Orden Recuperación	Suma Relevantes	Precision	Recall
1	1	1	1	0.2
1	2	2	1	0.4
1	3	3	1	0.6
0	4	3	0.75	0.6
1	5	4	0.8	0.8
1	6	5	0.83333333	1

*Tabla 75: Resultados consulta 1*

Esto quiere decir que de los 6 videos recuperados por el motor de búsqueda 5 de ellos son relevantes con la consulta realizada, el 6 video puede contener alguna palabra de las buscadas, pero realmente no es relevante con la consulta.



*Ilustración 26: Grafica Precision-Recall consulta 1*

Con esta grafica se puede apreciar la precisión del sistema para esta colección y esta consulta.

La segunda consulta realizada es: Graphic card expensive. Buscando que el sistema liste como resultados los videos incluidos en los que se habla de tarjetas gráficas, además de estos videos existen otros de educación que explican el cómo jugar al póker, por lo que también se hablara de cards, afectando a los resultados.



Relevancia	Orden Recuperación	Suma Relevantes	Precision	Recall
1	1	1	1	0.2
0	2	1	0.5	0.2
1	3	2	0.66666667	0.4
1	4	3	0.75	0.6
1	5	4	0.8	0.8
0	6	4	0.66666667	0.8
0	7	4	0.57142857	0.8
0	8	4	0.5	0.8
1	9	5	0.55555556	1
0	10	5	0.5	1

Tabla 76: Resultados consulta 2

En este caso se han obtenido 10 resultados, siendo únicamente relevantes 5 de ellos, por lo que la gráfica de Precision-Recall obtenida es la siguiente:

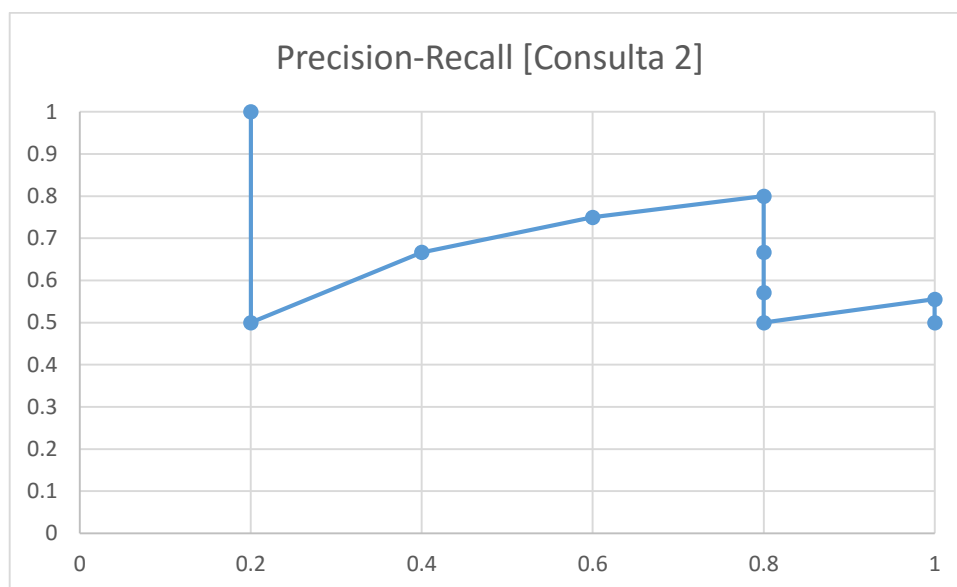


Ilustración 27: Gráfica Precision-Recall consulta 2

Apreciando como afecta la aparición de la palabra card en otros videos que no tienen que ver con lo que se busca.

Por último, se realiza la consulta: How to play with jacks. Con esta consulta se busca obtener en los resultados los videos relacionados con el póker, que expliquen cómo jugar cuando se tiene una o varias J. Los resultados obtenidos son:

Relevancia	Orden Recuperación	Suma Relevantes	Precision	Recall
1	1	1	1	0.14285714
1	2	2	1	0.28571429
1	3	3	1	0.42857143
1	4	4	1	0.57142857
0	5	4	0.8	0.57142857
1	6	5	0.83333333	0.71428571
1	7	6	0.85714286	0.85714286
0	8	6	0.75	0.85714286
0	9	6	0.66666667	0.85714286
1	10	7	0.7	1

Tabla 77: Resultados consulta 3

En este caso de los 10 videos recuperados 7 son relevantes y 3 no lo son, por lo que con esta consulta se obtienen resultados más precisos que con la anterior utilizando la colección de videos que posee el sistema.

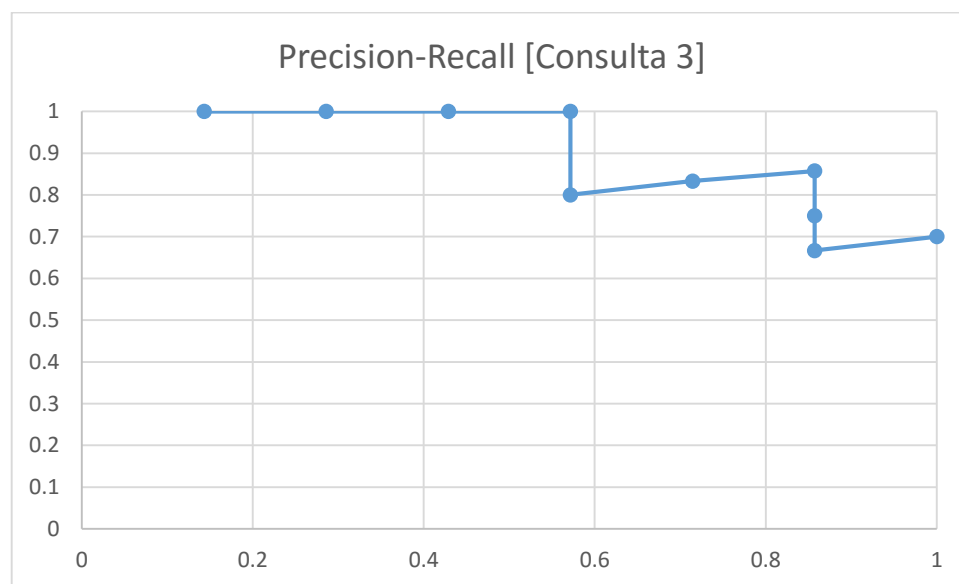


Ilustración 28: Grafica Precision-Recall consulta 3

Con estos cálculos de valoración del sistema no se pretende decir que es un sistema efectivo ni mucho menos, ya que el objetivo del proyecto era el plantear la idea creando un sistema que utilizara los subtítulos, y no se intenta meter en el campo de la mejora del motor de búsqueda, ya que apenas se ha modificado. Es por esto por lo que las consultas analizadas son consideradas como favorables para el sistema, ya que no tiene sentido el realizar búsquedas de contenido que se sabe que no existe en la colección de videos.

A pesar de que el sistema no mejorara el proceso de indexación y obtención de resultados se puede decir que tiene un buen funcionamiento con lo ofrecido por el motor de búsqueda Solr de forma nativa.

Las únicas modificaciones realizadas sobre el motor de búsquedas original son la de la inclusión de los campos necesarios para los videos y el de añadir la lista de Stopwords que limpian el contenido de las transcripciones de palabras que no aportarían valor a la hora de realizar consultas.

## 6. Implementación

En este capítulo se describirá como se ha llevado a cabo la implementación de la arquitectura y el diseño del sistema, que ha sido detallado en apartados anteriores. En esta ocasión, se mostrarán los aspectos más importantes de la fase de desarrollo tras el análisis y diseño del sistema, presentando las tecnologías utilizadas para la implementación del proyecto.

### 6.1. Tecnologías utilizadas

Tal como se avanzó, la aplicación será realizada mediante el lenguaje de programación Java, utilizando el entorno de desarrollo Eclipse. El sistema podrá ser ejecutado desde cualquier sistema operativo, mediante el uso de un navegador web.

Al tratarse de una página web que pretende ser utilizada en cualquier tipo de dispositivos, tal y como se indica en los requisitos, se busca que sea responsive, y para ello se deberá de utilizar el framework de Bootstrap, cuyo uso será brevemente explicado en este apartado.

Además, también se mostrarán los comandos necesarios para la creación de los índices sobre Solr y como añadir nuevos videos al sistema.

#### 6.1.1. Java Server Pages

En primera instancia, se procede a explicar la tecnología Java utilizada para generar contenido dinámico para la aplicación web, en forma de documentos HTML. Este conjunto de técnicas basadas en Java es denominado como Java Server Pages, también conocido por sus siglas, JSP.

Esta tecnología, permite la utilización de código Java mediante el uso de scripts, pudiendo ejecutar todo de forma similar a la que construye un Servlet, pero de una manera más simplificada.

El sistema de funcionamiento de esta tecnología se basa en que el encargado de interpretar el código contenido en la página JSP, el servidor web, en este caso Tomcat, que construye el código Java del Servlet que se debe generar. Y es este Servlet el encargado de llamar a los RequestHandler necesarios para que se encarguen de la petición y formen el documento que se presentará al usuario en el navegador web.

A continuación, se muestra un ejemplo de página JSP de este proyecto, detallando las tecnologías presentes en el mismo.

```
1  <%@ page import="java.util.*" %>
2
3  <!DOCTYPE html>
4  <html lang="es">
5  <head>
6  <title>
7  <%
8      if(request.getAttribute("Busqueda") != null){
9          %>
10         <%= request.getAttribute("Busqueda") %> - SubSearch
11         <%
12             }else{
13                 %>
14                 Inicio - SubSearch
15             <%
16                 }
17             %>
18         </title>
```

*Ilustración 29: Código ejemplo JSP*

Como se puede apreciar, los documentos JSP son documentos HTML con código Java que se encierra entre `<% y %>`, esto es conocido como scriptlet.

Las páginas que contiene el sistema son generadas dinámicamente a partir de una plantilla que contiene el código Java que permite recorrer los resultados que le ha pasado la clase RequestHandler, para así poder mostrar las listas de resultados obtenidos, la paginación de resultados, etc.

### 6.1.2. Solr

Del motor de búsquedas utilizado hay que destacar algún cambio que se ha realizado sobre los índices, además de explicar los comandos utilizados para crear los índices y añadir videos a los mismos.

Los cambios que se han realizado sobre los índices ha sido el añadirle una lista de stopwords o palabras vacías que no aportan valor a la hora de realizar búsquedas [31]. Esto se ha modificado cambiando el archivo de Stopwords que incluye Solr para cada Core por el obtenido del contenido de la página web referenciada.

Una vez ya se conocen los índices que se van a utilizar y los cambios que se han realizado sobre sus archivos de configuración, falta explicar la secuencia de comandos que hay que introducir para la creación de los mismos.

1. Descargar Solr y extraerlo en cualquier directorio.
2. Acceder al directorio mediante la consola.

```
[→ ~ cd Documents/Proyectos/solr-6.0.0
```

*Ilustración 30: Comando acceso al directorio Solr*

3. Arrancar el servidor de Solr.

```
[→ solr-6.0.0 bin/solr start
Waiting up to 30 seconds to see Solr running on port 8983 [/]
Started Solr server on port 8983 (pid=2255). Happy searching!
```

*Ilustración 31: Comando arranque servidor Solr*

4. Crea el core o índice de Solr donde se indexarán los documentos.

```
[→ solr-6.0.0 bin/solr create -c ejemplo

Copying configuration to new core instance directory:
/Users/Carlos/Documents/Proyectos/solr-6.0.0/server/solr/ejemplo

Creating new core 'ejemplo' using command:
http://localhost:8983/solr/admin/cores?action=CREATE&name=ejemplo&instanceDir=ejemplo

{
  "responseHeader":{
    "status":0,
    "QTime":363},
  "core":"ejemplo"}
```

*Ilustración 32: Comando creación Core Solr*

5. Parar el servidor web y realizar los cambios sobre los archivos de configuración del Core (Añadir Stopwords, añadir campos de videos).

```
➦ solr-6.0.0 bin/solr stop
Sending stop command to Solr running on port 8983 ... waiting 5 seconds to allow Jetty process 1923 to stop gracefully.
```

*Ilustración 33: Comando parar servidor Solr*

6. Arrancar el servidor Solr una vez realizados y guardados los cambios.
7. Indexar en el Core seleccionado los documentos que se encuentran en el directorio introducido como parámetro.

```
➦ solr-6.0.0 bin/post -c ejemplo -t ejemplo -l /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/*.*.xml
java -classpath /Users/Carlos/Desktop/Documentos/solr-6.0.0/dist/solr-core-6.0.0.jar -Dauto=yes -Dc=ejemplo -Ddata=files org.apache.solr.util.SimplePostTool /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/0Mdk8hBkR3s.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/1oosPQnxUJA.xml /Users/Carlos/Desktop/Documentos/2TtdPbeKNF
c.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/3zYnSL9oolw.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/455ytzjNSBc.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/6bwzNg5Q00e.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/7Xdc5XVUGe.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/BNHR61QJGZs.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/IZZ4y5GfDU.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/JQSTRKGE1Ww.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/LCoPLFaeq0U.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/MnbQjpejZt4.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/OTk40s6_Yc.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/RgKAFK5dJ5k.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/Tnt-MMAKXdk.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/V6zpM3p8R0.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/aMcK11T5Z2s.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/kn9ZWXCKr0M.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/l9cpHsr-gRk.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/tXVDY1HvrVU.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/ts9NDp518zA.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/vJZqTdMw53c.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/yaiPYu6Eduw.xml /Users/Carlos/Desktop/Documentos/SubSearch/Documentos/z1wYbVx_-gg.xml
SimplePostTool version 5.0.0
Posting files to [base] url http://localhost:8983/solr/ejemplo/update...
Entering auto mode. File endings considered are xml,json,jsonl,txt,doc,docx,ppt,pptx,xls,xlsx,odt,odp,ods,ott,otp,ots,rtf,htm,html,txt,log
Posting file 0Mdk8hBkR3s.xml (application/xml) to [base]
Posting file 1oosPQnxUJA.xml (application/xml) to [base]
Posting file 2TtdPbeKNF.xml (application/xml) to [base]
Posting file 3zYnSL9oolw.xml (application/xml) to [base]
Posting file 455ytzjNSBc.xml (application/xml) to [base]
Posting file 6bwzNg5Q00e.xml (application/xml) to [base]
Posting file 7Xdc5XVUGe.xml (application/xml) to [base]
Posting file BNHR61QJGZs.xml (application/xml) to [base]
Posting file IZZ4y5GfDU.xml (application/xml) to [base]
Posting file JQSTRKGE1Ww.xml (application/xml) to [base]
Posting file LCoPLFaeq0U.xml (application/xml) to [base]
Posting file MnbQjpejZt4.xml (application/xml) to [base]
Posting file OTk40s6_Yc.xml (application/xml) to [base]
Posting file RgKAFK5dJ5k.xml (application/xml) to [base]
Posting file Tnt-MMAKXdk.xml (application/xml) to [base]
Posting file V6zpM3p8R0.xml (application/xml) to [base]
Posting file aMcK11T5Z2s.xml (application/xml) to [base]
Posting file kn9ZWXCKr0M.xml (application/xml) to [base]
Posting file l9cpHsr-gRk.xml (application/xml) to [base]
Posting file tXVDY1HvrVU.xml (application/xml) to [base]
Posting file ts9NDp518zA.xml (application/xml) to [base]
Posting file vJZqTdMw53c.xml (application/xml) to [base]
Posting file yaiPYu6Eduw.xml (application/xml) to [base]
Posting file z1wYbVx_-gg.xml (application/xml) to [base]
24 files indexed.
COMMITTING Solr index changes to http://localhost:8983/solr/ejemplo/update...
Time spent: 0:00:00.772
```

*Ilustración 34: Comando indexación de documentos*

Estos son los pasos que hay que dar para añadir nuevos videos al índice, y así incluirlos en el sistema.

Al existir dos índices que indexan documentos que tienen distintos campos se ha decidido organizar estos documentos con la siguiente estructura de directorios:

- .
- /Documentos SubSearch/
  - o /Documentos/
    - Id\_video.xml
  - o /DocumentosMT/
    - /Id\_video/
      - Id\_video(XXX).xml (Siendo XXX un numero generado iterativamente).

### 6.1.3. Bootstrap

El Framework de desarrollo de la interfaz de usuario elegido fue Bootstrap en su versión 3.3.7 [32], ya que permite crear páginas web responsive, adaptándose a distintos tamaños de pantallas.

Para ello se ha de descargar sus hojas de estilos, funciones de JavaScript... que permiten colocar y escalar los elementos en pantalla. Algunas de las clases utilizadas son las siguientes:

```
<div class="row">
  <div class="col-sm-8 col-sm-offset-2 text-center">
    <h2>No se han encontrado resultados</h2>
  </div>
</div>
```

*Ilustración 35: Ejemplo uso clases Bootstrap*

Con el uso de las clases col- se puede colocar los elementos de una u otra forma dependiendo del tamaño de pantalla (que puede ser xs, sm, ms y lg), jugando con el tamaño de columna [33] y así llegar a crear una interfaz web como la del sistema.

## 6.2. Resultado final del sistema

A continuación, se mostrarán algunas capturas del resultado final del sistema, mostrando capturas realizadas desde distintos dispositivos para demostrar que la página se adapta al tamaño de pantalla del dispositivo donde se visualiza.

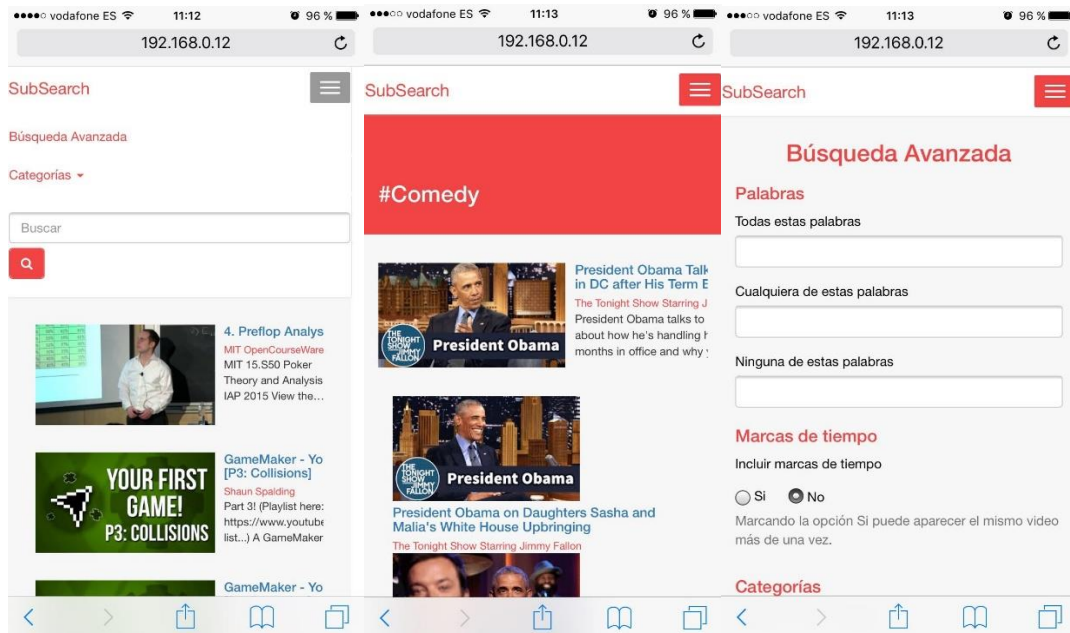


Ilustración 36: Ejemplos visualización dispositivo móvil

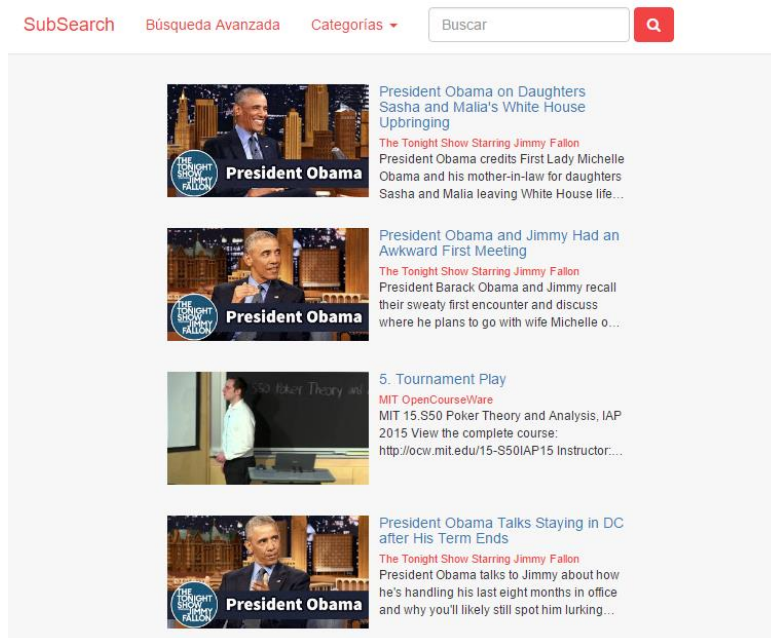


Ilustración 37: Ejemplo visualización pantallas grandes



## 7. Gestión del proyecto

A lo largo de este capítulo se detalla la planificación y la gestión del proyecto llevado a cabo a lo largo de su realización, comparando los tiempos estimados con los tiempos reales y finales. Por otra parte, se expondrá la estimación del presupuesto del proyecto.

### 7.1. Planificación temporal

Este apartado tiene como finalidad exponer la gestión y la planificación llevada a cabo, para lo cual se comparará la planificación realizada en un primer momento con la duración que verdaderamente ha llevado completar el proyecto.

#### 7.1.1. Planificación inicial

La fecha de comienzo del proyecto se establece el día 11 de abril de 2016, siendo la fecha estimada de finalización del proyecto el 30 de septiembre de 2016 (aproximadamente seis meses).

De forma estimada, el esfuerzo medio diario es de 4 horas al día de lunes a viernes, excluyendo fines de semana y festivos.

A continuación, se muestra la planificación inicial:

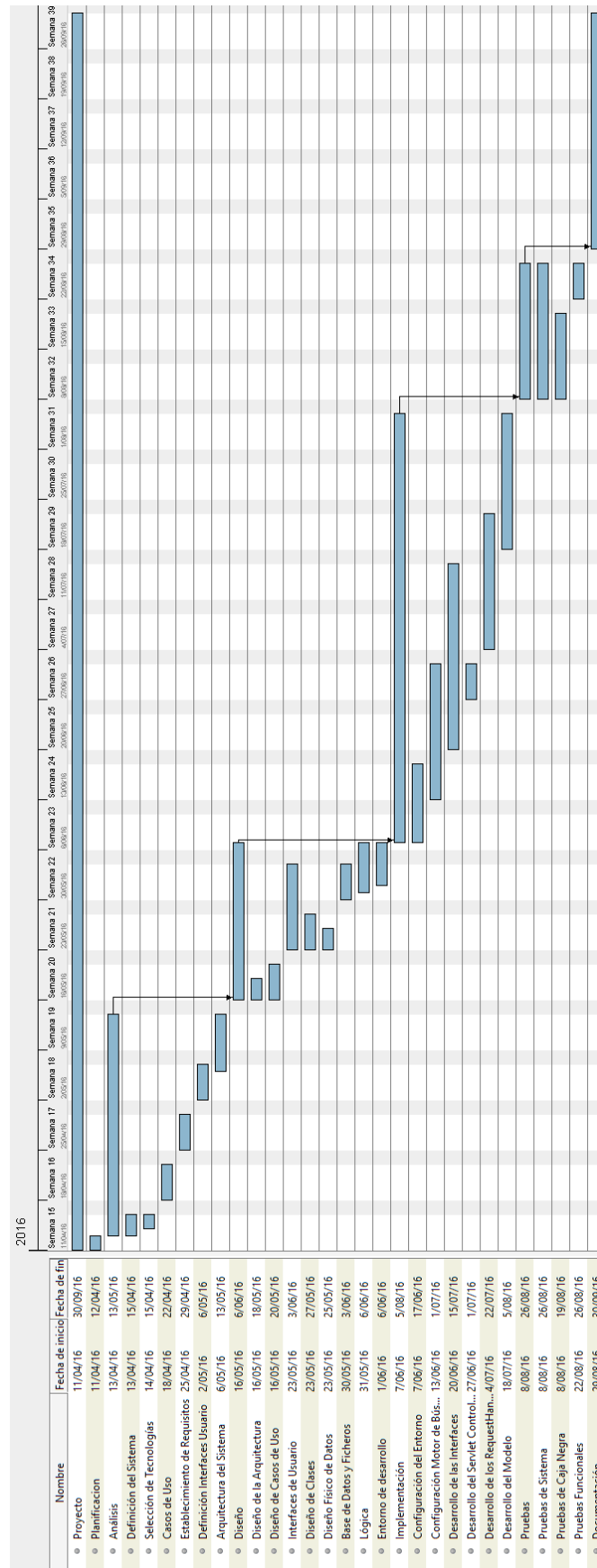


Ilustración 38: Planificación inicial estimada

### 7.1.2. Desarrollo real del Proyecto

Tras lo expuesto en el apartado anterior, la duración de algunas de las tareas expuestas ha variado con respecto a lo estimado en un primer momento. Esto se debe a los parones que se tuvieron que hacer por motivos de estudio a mediados de mayo, cuando se realizaban los exámenes del segundo cuatrimestre, además de los parones realizados en los días festivos del mes de agosto. Por esto es por lo que la estimación inicial es diferente a la real, además de la poca experiencia que se tiene a la hora de planificar y establecer periodos de tiempo en este tipo de proyectos.

Por lo tanto, a lo largo de este apartado se mostrarán las desviaciones que ha sufrido el proyecto, así como el tiempo total real que se ha dedicado para realizarlo completamente.

La fecha de inicio del proyecto fue el 11 de abril de 2016 y la fecha real de finalización fue el 23 de septiembre de 2016.

Como se puede observar en el diagrama de Gantt que ilustra el desarrollo real del proyecto, no sufrió ningún retraso, todo lo contrario, ya que la fecha de entrega se adelantó una semana. El no haberse sufrido retrasos se debe principalmente a la cantidad de días de los que se disponía para realizar el proyecto.

A pesar de no haberse cometido retrasos, no se ha cumplido la duración estimada de las actividades o tareas por los motivos comentados anteriormente, además de por las dificultades que han aparecido en algunos momentos de la fase de Implementación. El principal problema fue el poco conocimiento que se tenía de Solr, por lo que se tuvo que aprender el funcionamiento básico para así poder utilizarlo tal y como se buscaba.

Finalmente, la planificación real que siguió el proyecto fue la siguiente:

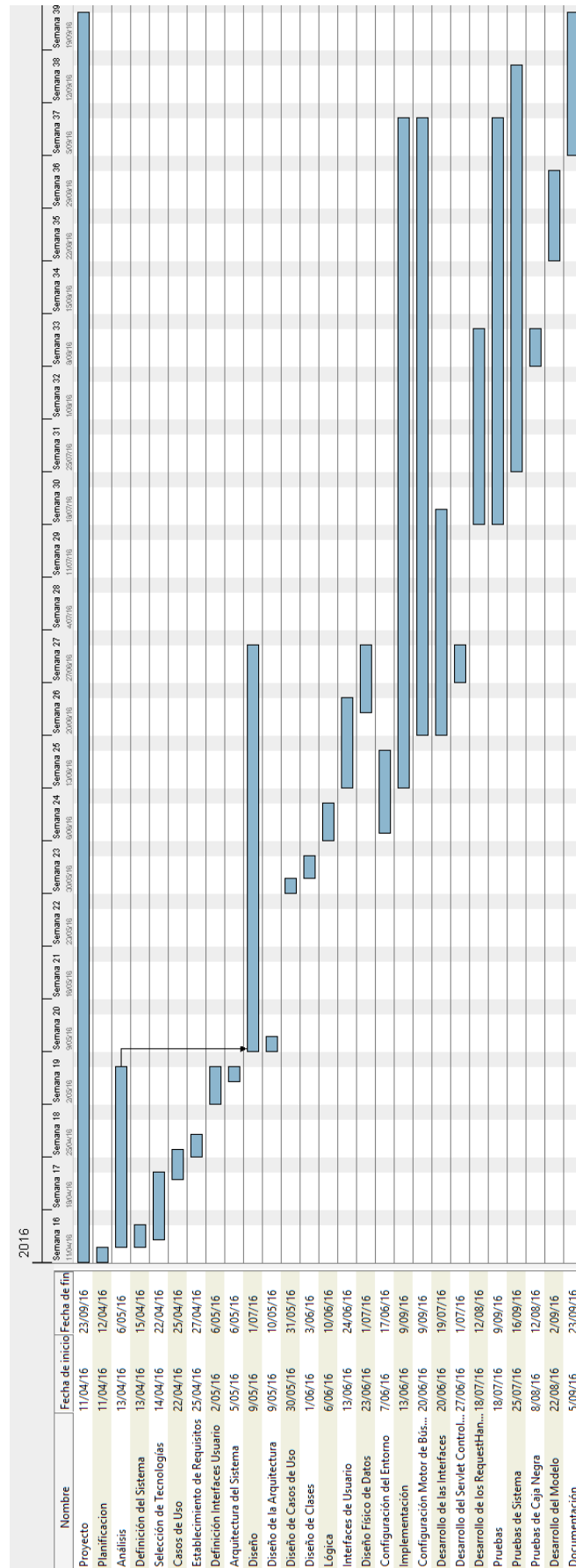


Ilustración 39: Desarrollo real del proyecto

Como se puede apreciar, existes varias tareas que se superponen las unas con las otras, sobre todo en la parte final del proyecto. Esto se debe a que se retocaban pequeñas partes conforme iba avanzando la implementación del sistema, además de acercarse la fecha de entrega del proyecto y esto suponer un mayor esfuerzo y dedicación.

Otro de los cambios que se aprecia respecto la estimación inicial es la eliminación de algunas tareas, ya que se consideraba que estaban cubiertas con otras tareas realizadas. Además, el cambio de duraciones de algunas actividades, que, al encontrarse con los parones de mayo y festivos de agosto, tuvieron que modificarse. Esto también se debe a que inicialmente se estimaba un esfuerzo diario de 4 horas, siendo el real más irregular, dedicando algunos días más tiempo y otros menos.

## 7.2. Presupuesto

Este apartado tiene como finalidad explicar y detallar el análisis económico del proyecto de forma estimada, basándose en la duración real del presente TFG. Para ello se presentará primero el coste total del proyecto, y en los siguientes apartados se desglosarán los costes.

### 7.2.1. Presupuesto total

El proyecto con título: *Sistema avanzado para la indexación y búsqueda de recursos audiovisuales basado en metadatos y contenido*. Con una duración total de **5 meses y 12 días** (incluyendo días no laborables).

Tiene un presupuesto total para su realización de **22,829.97€ (VEINTIDOS MIL OCHOCIENTOS VEINTINUEVE CON NOVENTA Y SIETE EUROS)** (I.V.A. incluido).

En los siguientes apartados se realizará un desglose del mismo en el que se justificaran cada uno de los costes.

### 7.2.2. Desglose presupuestario

En este apartado se va a indicar cada uno de los gastos que se han tenido en cuenta a la hora de calcular el presupuesto total del proyecto.

#### 7.2.2.1 Costes directos

##### *Costes de Personal*

Durante la realización del proyecto, el alumno ha tenido que realizar tareas que habitualmente son realizadas por distintos roles dentro del desarrollo de un proyecto, es por esto por lo que se diferenciaron los siguientes roles:

- **Jefe de Proyecto:** encargado de supervisar el trabajo del resto de personas y asegurarse de que el proyecto cumple con los objetivos marcados.
- **Analista:** encargado de analizar las necesidades del usuario y plasmarlas en casos de uso, requisitos y elección de tecnologías, entre otros.
- **Diseñador:** encargado de diseñar el sistema en función de lo analizado por el analista. Diseñará las clases, el modelo físico de datos, las interfaces, etc.
- **Programador:** encargado de llevar a cabo lo diseñado por el diseñador.
- **Gestor de Pruebas:** encargado de comprobar que todo lo implementado por el programador funciona correctamente y cumple las necesidades acordadas con el cliente.

Estos roles al realizar diferentes funciones dentro del proyecto recibirán salarios distintos, que serán calculados utilizando la información de la guía de mercado laboral del año 2016 [34], siendo la experiencia de cada rol de 0-2 años.

Rol	Horas Dedicadas (h)	Coste (€/h)	Coste
Jefe de Proyecto	100	38.05	3,804.71 €
Analista	64	26.40	1,689.60 €
Diseñador	92	25.24	2,321.65 €
Programador	250	21.74	5,435.29 €
Gestor de Pruebas	80	18.64	1,490.82 €
TOTAL	586	-	14,742.07 €

*Tabla 78: Costes de personal*

### **Costes de material**

El proyecto también ha conllevado el gasto de material informático (Hardware y Software) y material fungible. Como los equipos utilizados ya eran propiedad del autor antes del comienzo del proyecto se realizará la amortización correspondiente a la duración del proyecto, suponiendo que el equipo informático se amortizará completamente en 3 años.

Unidades	Componente	Coste	Meses de uso	Coste aplicable al proyecto
1	Apple MacBook Pro	1,500.00 €	6	250.00 €
1	Ordenador sobremesa	800.00 €	6	133.33 €
1	Monitor HP 23"	190.00 €	6	31.67 €
1	iPhone 6s 16GB	550.00 €	6	91.67 €
1	Licencia Windows 10	280.00 €	6	46.67 €
1	Microsoft Office 2016	150.00 €	6	25.00 €
1	Licencia Sublime Text	70.00 €	6	11.67 €
TOTAL		590.00 €		

Tabla 79: Costes de material informático

Unidades	Artículo	Coste unidad	Coste
1	Paquete Papel Din A-4	3.75 €	3.75 €
3	Bolígrafo BIC	0.27 €	0.81 €
1	Paquete 100 Post-it	3.00 €	3.00 €
TOTAL		7.56 €	

Tabla 80: Costes de material fungible

#### **7.2.2.2 Costes indirectos**

Los costes indirectos asociados al proyecto se calcularán mediante una estimación porcentual sobre los costes directos, estableciendo el porcentaje en 5%. En estos gastos indirectos se incluyen gastos como el de luz, agua, conexión a internet, teléfono, etc.

Costes Indirectos	
5% sobre gastos Directos	766.98 €

Tabla 81: Costes indirectos

El beneficio a obtener para la realización de este proyecto ha sido estimado basándose en todos los estudios y datos obtenidos para el mismo. En este estudio se ha concretado un beneficio del 18% sobre los costes directos del proyecto.

Beneficios	
18% sobre gastos Directos	2,761.13 €

Tabla 82: Beneficios

### 7.2.3. Resumen de costes

Finalmente, el resumen de costes mostrara los costes totales del proyecto sin IVA y con IVA, agrupando los costes que se han ido anunciando en apartados anteriores.

Resumen de Costes	
Costes Directos	
Salarios	14,742.07 €
Material informático	590.00 €
Material fungible	7.56 €
Costes indirectos	766.98 €
Beneficios	2,761.13 €
TOTAL sin IVA	18,867.75 €
IVA (21%)	3,962.23 €
TOTAL con IVA	22,829.97 €

*Tabla 83: Resumen de costes*



## 8. Marco regulador y ético

Este apartado se dedica al aspecto legal de todo lo referente a este proyecto. Se detallarán los distintos estándares o licencias de aquellas tecnologías utilizadas en el proyecto

### 8.1. Java

Al tratarse del lenguaje de programación utilizado para el desarrollo del sistema, se ha de remarcar que su compañía, Oracle, concede la licencia de software siempre y cuando se acepten cada uno de los términos de su licencia [35], ya que se trata de un lenguaje de programación libre o abierto. Por lo que básicamente se han aceptado los términos de licencia.

### 8.2. Solr

Solr al pertenecer a la Apache Software Foundation utiliza las licencias de la misma. La ASF utiliza una licencia de software libre permisiva, que requiere la preservación de copyright y exención de responsabilidad. Permitiendo al usuario del software la libertad de usarlo para cualquier propósito, para distribuirlo, modificarlo y distribuir versiones modificadas del software, bajo los términos de la licencia, sin preocuparse.

Por este motivo, se ha añadido un documento al proyecto con la siguiente licencia que se puede conseguir desde su página web [36]:

Copyright 2016 Carlos Contreras Sanz

Licensed under the Apache License, Version 2.0 (the "License");  
you may not use this file except in compliance with the License.

You may obtain a copy of the License at

<http://www.apache.org/licenses/LICENSE-2.0>

Unless required by applicable law or agreed to in writing, software  
distributed under the License is distributed on an "AS IS" BASIS,  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.

See the License for the specific language governing permissions and  
limitations under the License.

### 8.3. Bootstrap

Con Bootstrap ocurre lo mismo, es de uso libre, pero para ello se ha de incluir su declaración de copyright, por lo que se ha incluido el siguiente documento:

The MIT License (MIT)

Copyright (c) 2011-2016 Twitter, Inc.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

### 8.4. YouTube

YouTube es otro de los pilares del sistema, por lo que al utilizar información del mismo se deberán de leer los términos y condiciones del servicio y así asegurar que no existe ningún problema legal.

Como se puede observar en la declaración de Términos y Condiciones que establece YouTube [37], permite acceder y utilizar los Servicios, con sujeción a las condiciones que imponen, que tras haber sido leídas se ha comprobado que el uso de YouTube que hace el sistema está bajo la legalidad.

## 9. Conclusiones y Trabajos Futuros

A lo largo de este apartado se procede a mostrar las conclusiones extraídas de la realización del presente trabajo fin de grado. Así mismo, se plantearán posibles futuras líneas de desarrollo que mejoren la funcionalidad del sistema.

### 9.1. Conclusiones

En primer lugar, se puede extraer como conclusión que las funcionalidades que se han implementado en el sistema se pueden considerar innovadoras, ya que aún no han sido utilizadas por ningún otro buscador, o al menos no se conoce ninguno que lo haga.

Las funcionalidades pensadas en un principio han sido cumplidas, con unos resultados que pueden llegar a dar de pensar que si una plataforma de video grande lo explotara podría llegar a dar frutos. Al igual que si surgiera una plataforma de videos nueva que utilizara rasgos de este sistema, pero siempre debería de ser aplicado a una plataforma y no seguir el ejemplo de este sistema que se apoya de YouTube. Esto se debe básicamente a la obtención de beneficios.

Dejando beneficios a un lado, el sistema se basa en dos pilares que con los avances de las tecnologías irán mejorando con el tiempo, y son el del reconocimiento de voz, ya que es necesario para poder conseguir las transcripciones de los videos y audios, y así enriquecer los metadatos de los mismos. El segundo pilar es el de las búsquedas de contenido multimedia, ya que como se adelanta en la introducción cada día se generan miles de horas de contenido, necesitando así mejorar los motores de búsquedas para poder servir lo mejor posible las consultas de los usuarios con el paso del tiempo e incremento del número de videos en este caso.

Por lo que la idea que pretende transmitir este trabajo puede ser que sea utilizada en el futuro por los motores de búsqueda de contenido multimedia una vez se encuentre una forma efectiva de procesar la información que pueden ofrecer las transcripciones de los videos.

## 9.2. Conclusiones personales

A modo de conclusión personal, no puedo si no alegrarme por el resultado obtenido y del duro trabajo realizado, que me ha sido útil para poder vivir en pequeña escala la cantidad de trabajo que conlleva el desarrollo de un sistema o aplicación software, en definitiva, el cómo se lleva a cabo un proyecto.

Todo esto ha sido posible gracias a los conocimientos que he ido adquiriendo como alumno durante estos últimos 4 años en asignaturas como Ingeniería del Software, Interfaces de Usuario, Recuperación y Acceso a la Información, Dirección de Proyectos y Desarrollo del Software, entre muchas otras. Que me han enseñado a adaptarme al uso de diferentes tecnologías, y a aprender nuevas tecnologías más rápidamente.

En definitiva y a modo de conclusión personal, he de reconocer que dependiendo del día y del estado de ánimo, me sentía o no motivado, y pienso que esto ha sido debido a que la idea con la que surge el proyecto pretende ser innovadora, y como tal no tienes con quien compararte, ya que la primera comparación que se puede realizar es con YouTube, que al ser la plataforma más grande de video que existe en internet es difícil de superar aun disponiendo de una selección de videos que no supone ni un 1% de la colección que posee YouTube. A pesar de esto pienso que si alguien consigue explotarlo y sacarle el máximo partido posible puede ser un buen método de búsquedas en el futuro.

### 9.3. Trabajos Futuros

Todo proyecto software es mejorable y este, no es una excepción. Por eso se pueden proponer varias mejoras que mejorarían el rendimiento del sistema, o que aportan funcionalidades que se buscaban ofrecer, pero que por falta de tiempo y falta de conocimientos no se han podido aplicar.

1. Como primera mejora, está la de plantear una plataforma de video independiente de YouTube que albergara sus propios videos, y que pueda crecer gracias a los usuarios, añadiendo ellos mismos sus videos. Encargándose el sistema de procesar de manera automática esos videos para enriquecer los metadatos con las transcripciones de los mismos.
2. Mejorar el motor de búsquedas agregándole el reconocimiento de entidades, entre otras mejoras, es decir, aplicar técnicas de procesamiento del lenguaje natural, lematización, tokenización...
3. Modificación del sistema para que utilice la API de YouTube y así poder llegar a agregar videos al sistema utilizando únicamente el enlace del video que se desea de agregar al índice.

## 10. Referencias

- [1] Ditrendia, «Informe Mobile en España y en el Mundo 2015,» 2015.
- [2] I. F. Lantigua, «El móvil supera por primera vez al ordenador para acceder a Internet,» *El Mundo*, 04 Abril 2016.
- [3] InternetLiveStats, «Internet Live Stats,» 15 Septiembre 2016. [En línea]. Available: <http://www.internetlivestats.com/internet-users/>. [Último acceso: 15 Septiembre 2016].
- [4] PC World, «PC World Mexico,» 22 Febrero 2012. [En línea]. Available: <http://www.pcworld.com.mx/Articulos/21539.htm>. [Último acceso: 15 Septiembre 2016].
- [5] DOMO, «Data Never Sleeps 4.0,» 28 Junio 2016. [En línea]. Available: <https://www.domo.com/blog/2016/06/data-never-sleeps-4-0/>. [Último acceso: 1 Septiembre 2016].
- [6] YouTube, «Subtitulos automaticos,» [En línea]. Available: [https://support.google.com/youtube/answer/6373554?hl=es&ref\\_topic=3014331](https://support.google.com/youtube/answer/6373554?hl=es&ref_topic=3014331).
- [7] IBM, «Speech to Text,» Watson Developer Cloud, [En línea]. Available: <https://www.ibm.com/watson/developercloud/speech-to-text.html>.
- [8] Alexa, «Alexa,» Amazon Company, 15 Septiembre 2016. [En línea]. Available: <http://www.alexa.com/topsites>. [Último acceso: 15 Septiembre 2016].
- [9] J. A. Muñoz, «InboundCycle,» 14 Noviembre 2014. [En línea]. Available: <http://www.inboundcycle.com/blog-de-inbound-marketing/gestion-metadatos-youtube-y-otros-secretos>. [Último acceso: 15 Septiembre 2016].
- [10] Apache Software Foundation, «Lucene,» 29 octubre 2014. [En línea]. Available: <https://lucene.apache.org/core/>. [Último acceso: 14 septiembre 2016].
- [11] The Apache Software Foundation, «Solr,» 2016. [En línea]. Available: <http://lucene.apache.org/solr/>. [Último acceso: 17 Septiembre 2016].
- [12] Apache Software Foundation, «Elasticsearch,» 2016. [En línea]. [Último acceso: 14 septiembre 2016].
- [13] Ministerio de Hacienda y Administraciones Públicas, Gobierno de España, «Metodología de Planificación, Desarrollo y Mantenimiento de sistemas de información,» 2001. [En línea]. Available:

[http://administracionelectronica.gob.es/pae\\_Home/pae\\_Documentacion/pae\\_Metodolog/pae\\_Metrica\\_v3.html#.V9cCFZOLSRs](http://administracionelectronica.gob.es/pae_Home/pae_Documentacion/pae_Metodolog/pae_Metrica_v3.html#.V9cCFZOLSRs). [Último acceso: 09 Septiembre 2016].

- [14] T. Zarraonandia Ayo y J. Hernando Corrochano, «MVC, Model-View-Controller. Tecnologías informáticas de la Web,» Madrid, 2013.
- [15] M. Domínguez-Dorado, Todo Programación, Madrid, Madrid: Editorial Iberprensa, 2015.
- [16] C. Targett, «Solr Wiki,» Solr, 20 08 2013. [En línea]. Available: <https://wiki.apache.org/solr/Solrj>. [Último acceso: 2016].
- [17] Y. ". Matsumoto, «informIT,» 12 Junio 2000. [En línea]. Available: <http://www.informit.com/articles/article.aspx?p=18225>. [Último acceso: 17 Septiembre 2016].
- [18] Ecma International, «Standard ECMA-262. ECMAScript® 2016, Language Specification,» Junio, Genova, 2016.
- [19] Eclipse, «Eclipse IDE for Java EE Developers,» 24 Junio 2015. [En línea]. Available: <https://eclipse.org/downloads/packages/eclipse-ide-java-ee-developers/mars2>. [Último acceso: 4 Septiembre 2016].
- [20] Apple Inc., «Xcode,» 3 Mayo 2016. [En línea]. Available: <https://developer.apple.com/xcode/>. [Último acceso: 6 Septiembre 2016].
- [21] D. Ho, «Notepad++,» [En línea]. Available: <https://notepad-plus-plus.org/>. [Último acceso: 14 Septiembre 2016].
- [22] J. Skinner, «Sublime Text,» 9 Febrero 2016. [En línea]. Available: <https://www.sublimetext.com/>. [Último acceso: 14 Septiembre 2016].
- [23] Google2SRT, «Google2SRT,» 6 Marzo 2016. [En línea]. Available: <http://google2srt.sourceforge.net/es/>. [Último acceso: 14 Septiembre 2016].
- [24] DownSub, «DownSub,» 11 Septiembre 2015. [En línea]. Available: <http://downsub.com/>. [Último acceso: 14 Septiembre 2016].
- [25] Google, «Google Chrome,» 31 Agosto 2016. [En línea]. Available: <https://www.google.com/chrome/browser/desktop/index.html>. [Último acceso: 14 Septiembre 2016].

- [26] Apple Inc., «Safari,» 21 Octubre 2015. [En línea]. Available: <http://www.apple.com/es/safari/>. [Último acceso: 14 Septiembre 2016].
- [27] Fundación Mozilla, «Mozilla Firefox,» 24 Agosto 2016. [En línea]. Available: <https://www.mozilla.org/es-ES/firefox/new/>. [Último acceso: 14 Septiembre 2016].
- [28] Fundación Mozilla, «Firebug,» 23 Septiembre 2011. [En línea]. Available: <http://getfirebug.com/>. [Último acceso: 14 Septiembre 2016].
- [29] World Wide Web Consortium, «W3C Markup Validation Service,» W3C, Noviembre 2008. [En línea]. Available: <https://validator.w3.org/>. [Último acceso: 14 Septiembre 2016].
- [30] GitHub, «GitHub Desktop,» 6 Septiembre 2016. [En línea]. Available: <https://desktop.github.com/>. [Último acceso: 14 Septiembre 2016].
- [31] Ranks NL, «English Stopwords,» [En línea]. Available: <http://www.ranks.nl/stopwords>. [Último acceso: 1 agosto 2016].
- [32] Bootstrap, «Bootstrap,» [En línea]. Available: <http://getbootstrap.com/>. [Último acceso: 15 junio 2016].
- [33] W3Schools, «W3Schools,» [En línea]. Available: [http://www.w3schools.com/bootstrap/bootstrap\\_grid\\_basic.asp](http://www.w3schools.com/bootstrap/bootstrap_grid_basic.asp). [Último acceso: 21 septiembre 2016].
- [34] HAYS, «Guía del mercado laboral 2016,» Hays, Barcelona, 2016.
- [35] Oracle America Inc., «ORACLE,» 2 abril 2013. [En línea]. Available: <http://www.oracle.com/technetwork/java/javase/terms/license/index.html>. [Último acceso: 21 septiembre 2016].
- [36] Apache Software Foundation, «License v2.0,» 10 enero 2014. [En línea]. Available: <http://www.apache.org/licenses/LICENSE-2.0>. [Último acceso: 21 septiembre 2016].
- [37] YouTube, «YouTube,» 9 junio 2010. [En línea]. Available: <https://www.youtube.com/t/terms>. [Último acceso: 21 septiembre 2016].



## Anexo A: Summary

### 1. Introduction

Each day Information Technologies and communication have more influence on society. That's mainly because smartphone use is growing every day, for example, in Spain 87% mobile phones are Smartphones [1], exceeding even internet accesses from personal computers [2].

This influence is also explained because of how easy it is to connect to the internet, and how spread the internet is. There are approximately three and a half billion internet users in the world nowadays, considering as a user any individual who has internet connection on his home [3].

Studies have demonstrated that everyday large amounts of data are generated by users on social networks like Twitter, YouTube, etc. Because of this the necessity for file compression, document searches, etc. appears.

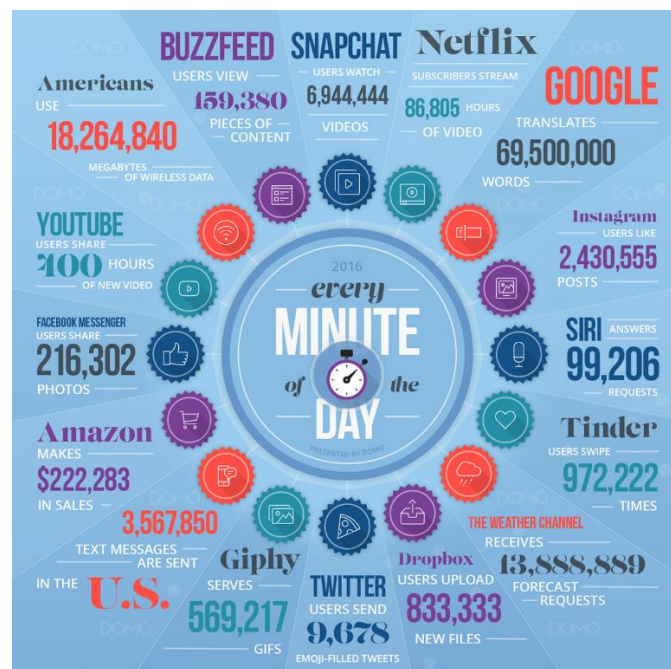


Illustration 40: Influence social networks per minute

This project aims to solve one of those problems, multimedia searches, in this case video searches, creating a webpage with a search engine that searches for videos using metadata adding more info than the usual. This new info are subtitles or transcriptions of the video.

## 1.1. Motivation

As seen in the introduction, there are services such as social networks or video platforms that create and consume a lot of information per minute, for example, in that minute, YouTube users share 400 hours of video [5], another example is Netflix, which users consume 86.805 hours of video.

That's why these platforms, which have such amount of information available to their users, aim to offer quality search services, with relevant results. But when the moment of truth comes, when the user performs multiple searches he realises that the query only returns results where the title or the description contain the words from the search box, in addition to that the results are affected by other factors such as the number of reproductions or amount of "Likes" received.

That's why it would be desirable to offer a service that doesn't take into consideration if the keywords appear in the title or description, but instead looking for keywords that appear in the real content, i.e. in the video itself. This was something impossible to achieve a few years ago, because there weren't the necessary technological tools, but nowadays there are voice recognition systems that have a very low error margin, for instance, YouTube's automatic subtitles [6], or Speech to Text tools as that of IBM [7].

In addition to adding this functionality to the world of online videos, it could be equally used in the world of radio or podcast, because it would work with the same content.

## 1.2. Objectives

As has been commented in the preceding paragraphs, the objective of this Final Project is designing a website that searches for videos using his own enhanced search engine.

Specifically, this project aims to analyse, design and develop the functions necessary for a website to be capable of using a search engine that will have indexed the information of them videos. So the system will consult the search engine, retrieving results and presenting them to the user.

In addition to this, the project also seeks to make it possible increase the number of videos offered by the system, so it provides the function to create the necessary documents for indexing new videos.

Another stated goal is to develop a small application to perform the conversion of an audio file to text, using any existing system Speech to Text, to use it when adding new videos to the index.

## 2. State of the art

This section seeks to outline the environment that would be closer to the system, analysing the different available alternatives that exist to carry out the idea of the system to be developed, and decisions taken. For example, the system will operate with the following characteristics:

- YouTube as a video platform providing videos to the system.
- IBM Speech to Text and YouTube Auto-subtitles as voice recognition systems.
- Solr as a search engine

### 3. System analysis

This section tries to get a detailed specification information system that meets the information needs of users and provide a basis for the subsequent design of the system. Briefly describing the proposed system, defining its scope and identify users who take part in the system. Use cases and requirements that have to meet the system will be defined. And finally it was found that all requirements complete the use cases and vice versa.

### 4. System Design

Throughout this chapter a design will be described in which the system architecture will be defined. In addition to this, the technological environment that will shape the system will be defined, and the different components of the system will be specified with more detail.

First, the details of each of the components that make up the system and its functions, class design, physical design data, user interfaces, the structure of files and databases will be explained.

### 5. Test plan

This section, tries to verify that the system is complying with the specified features in sections 3 and 4, System Analysis and Design.

Thus, for the development of the different test, the already specified requirements, use cases, architecture ... will be considered.

Before specifying system tests, the scope and test environment will be defined, to know what it's going to be tested and the environment it is going to be tested in.

## 6. Implementation

This chapter will describe how the implementation of the architecture and design of the system were carried out, which has been detailed in previous sections. This time, the most important aspects of the development phase will be shown after analysis and system design, introducing the technologies used for the project implementation.

## 7. Project management

Throughout this chapter the planning and project management carried out along the project are detailed, comparing the estimated times with the actual end times. Moreover, the estimate of the project budget will be discussed.

## 8. Regulatory and ethical framework

This section is dedicated to the legal aspect of all about this project. Different standards or licensing of the technologies used:

- Java
- Bootstrap
- Solr

## 9. Conclusions and Future Lines of Research

During this section it will be shown the main conclusions deduced of this TFG. Likewise, it will be displayed the possible future lines of investigation that will increase the system's usefulness and possible improvements.

### 9.1. Conclusions

First of all, we can deduce as conclusion that the functionalities that have been implemented in the system can be considered innovative, as they haven't been used by any other search engine, or at least none is known to do it.

The functionalities have been fulfilled as initially designed, with results that can lead to believe that if a large video platform exploits this functionality could bear fruit. It could also be very beneficial for any new video platforms using features of this system, but it should always be applied to a platform and not follow the example of this system based on YouTube. This is primarily due to the profit.

Leaving benefits aside, the system is based in two pillars that with advances in technology will improve over time which are the speech recognition, as it is necessary to get the transcripts of the videos and audios to enrich metadata. The second pillar is the search of multimedia content, as it is advanced in the introduction each day are generated thousands of hours of content, therefore needing to improve search engines to serve user queries with the increasing number of videos.

So the idea that this project tries to convey is that this system can be used in the future by search engines of multimedia content when someone finds a way to process the information provided by the transcriptions of the videos.

## 9.2. Personal Conclusions

As a personal conclusion, I cannot but be pleased by the hard work done, which has been useful for me, living on a small scale the amount of work involved in the development of a system or application, in other words, how a project is carried out.

All this has been possible thanks to the knowledge that I have gained as a student during the last 4 years in subjects like Software Engineering, User Interfaces, Recovery and Access to Information, Project Management and Software Development, among many others. I have been taught to adapt to the use of different technologies, and how to learn new technologies faster.

In short and making a personal conclusion, I must admit that depending on the day and mood, I felt motivated or not. I think that was mainly because the idea behind the project is to be innovative, which is a great thing, but for that same reason I can't really compare the system with others. One of the few comparisons that could be made is with YouTube, which is the largest video platform that exists on the Internet, and it is hard to beat its search results even working with a selection of videos that includes less than 1% of the collection hosted on YouTube. Despite this I am convinced that if someone manages to exploit and get the most out of this idea, it can be a good method for future searches.

## 9.3 Future work

All software can be improved and this project is no exception. Developers taking on the development of this idea may propose several enhancements that improve system performance, or provide new features, but for lack of time and lack of knowledge they have not been applied.

1. A first improvement could be to create an independent video platform from YouTube, hosting its own videos, which can grow thanks to users themselves adding their videos. The system would automatically process the videos and would enrich them with the transcriptions metadata.
2. Improve search engine adding entity recognition, among other improvements, i.e. applying techniques of natural language processing, stemming, tokenization...
3. Changing the system to use the YouTube API so you can get to add video to the system using only the video link you want to add to the index.



## Anexo B: Colección de videos

Para la realización de las pruebas del sistema, tanto de rendimiento como de funcionalidad se ha utilizado una colección de 50 videos escogidos al azar, tras comprobar que los subtítulos eran de calidad. Los videos se han seleccionado por categorías, existiendo cuatro (Educación, Entretenimiento, Comedia y Tecnología) teniendo como idiomas el inglés y español. En todo momento se conoce cada uno de los videos que forma el sistema, por lo que al realizar búsquedas concretas se sabrá si los resultados obtenidos son relevantes o no.

A continuación, se listarán los enlaces a los videos que forman la muestra:

### ***EDUCACION (#Education)***

[https://www.youtube.com/watch?v=OTkq4OsG\\_Yc](https://www.youtube.com/watch?v=OTkq4OsG_Yc)  
<https://www.youtube.com/watch?v=tXVDY1HvrVU>  
<https://www.youtube.com/watch?v=LCoPLFaeg0U>  
<https://www.youtube.com/watch?v=tssNDp5l6zA>  
<https://www.youtube.com/watch?v=JQSTRkGEiWw>  
<https://www.youtube.com/watch?v=IZZ4y5GfdOU>  
<https://www.youtube.com/watch?v=MnbQipejZt4>  
<https://www.youtube.com/watch?v=kn92WXcKr0M>  
<https://www.youtube.com/watch?v=3zYNSL9Oolw>  
<https://www.youtube.com/watch?v=6bwzNg5qQ0o>  
<https://www.youtube.com/watch?v=7XDcSXVUGsE>  
<https://www.youtube.com/watch?v=455ytzjNSBc>  
<https://www.youtube.com/watch?v=Tnt-MMAkXDk>  
<https://www.youtube.com/watch?v=l9cpHsr-gRk>  
<https://www.youtube.com/watch?v=1oosPQnxUjA>  
<https://www.youtube.com/watch?v=3Q3eITC01Fg>  
<https://www.youtube.com/watch?v=TnzcwTyr6cE>

### ***ENTRETENIMIENTO (#Entertainment)***

<https://www.youtube.com/watch?v=0MdK8hBkR3s>  
<https://www.youtube.com/watch?v=V6zpM3pm8rU>  
<https://www.youtube.com/watch?v=bWtZewDHVOw>  
<https://www.youtube.com/watch?v=AdseldK2Ga8>  
<https://www.youtube.com/watch?v=ALNKF0Lzew4>  
<https://www.youtube.com/watch?v=bQL3QDzh4ns>  
<https://www.youtube.com/watch?v=qMYt3dhxjDU>

<https://www.youtube.com/watch?v=OA7ZAJuKbOc>  
<https://www.youtube.com/watch?v=-9EcHaL6f6g>  
<https://www.youtube.com/watch?v=8EERYW5KX-U>

### **COMEDIA (#Comedy)**

[https://www.youtube.com/watch?v=ziwYbVx\\_-qg](https://www.youtube.com/watch?v=ziwYbVx_-qg)  
<https://www.youtube.com/watch?v=aMcKi1TS2Zs>  
<https://www.youtube.com/watch?v=iL-cP2mDPTA>  
<https://www.youtube.com/watch?v=yaiPYu6Eduw>  
<https://www.youtube.com/watch?v=vJZqTdWNd3c>  
<https://www.youtube.com/watch?v=2TtdPbeKNFc>  
<https://www.youtube.com/watch?v=wSV1UFbhJ-A>  
<https://www.youtube.com/watch?v=-8DwOOCJYzA>  
<https://www.youtube.com/watch?v=5SH7KODhiuY>  
<https://www.youtube.com/watch?v=ABRNKbSw-30>

### **TECNOLOGIA (#Technology)**

<https://www.youtube.com/watch?v=BNHR6IQJGZs>  
<https://www.youtube.com/watch?v=wMWBMDqKt8>  
<https://www.youtube.com/watch?v=pVidh1OZvHs>  
<https://www.youtube.com/watch?v=TCaOwDsY8T4>  
<https://www.youtube.com/watch?v=jpjlmpDTgCg>  
<https://www.youtube.com/watch?v=KDewTMxHw1M>  
[https://www.youtube.com/watch?v=ACFaeAo\\_jUg](https://www.youtube.com/watch?v=ACFaeAo_jUg)  
<https://www.youtube.com/watch?v=BXvRIbdMEr4>  
<https://www.youtube.com/watch?v=JMSq8nmDT00>  
[https://www.youtube.com/watch?v=YK\\_vHBwWfqs](https://www.youtube.com/watch?v=YK_vHBwWfqs)  
[https://www.youtube.com/watch?v=r0fgEVEgK\\_k](https://www.youtube.com/watch?v=r0fgEVEgK_k)

